Verifiability Structure, Collusion and Informativeness

with an application to the reform of contracting system for Chinese state enterprises

Thesis submitted

in partial fulfilment of the requirements

for the degree of

Master of Philosophy in Economics

by

Kim-Sau Chung[1]

Department of Economics

The Chinese University of Hong Kong

Shatin, Hong Kong

June 1994

1 I wish to thank Kai-Yuen Tsui and Wing-Chung Pun for useful discussion. I also want to thank my classmates, especially Wai-Hin Yan, Wai-Kin Lai, Chun-Fai Lau and Kwok-Fai Chung. Their spiritual support is indispensible to the eventual finish of this study. Last but not least, I would like to thank Yuk-Shing Cheng. He teaches me what should a Chinese student of economics study, from which I hope I will not deviate much in the rest of my academic life.

Abstract

We study the class of principal-informer-agent problems that includes decision making problems under uncertainty, principal-agent problems with hidden information, principal-supervisor-agent problems, and many others. We introduce the concept of verifiability structure which plays an important role in determining the "informativeness" of an informer. With the help of the concept of verifiability structure, we derive some preliminary insights on the "informativeness" of informers. We also apply our findings to the Chinese central government's problem of signing responsibility contracts with its state enterprises, and show why adopting idiosyncratic responsibility contracts may be undesirable in many cases.

Table of Content

Table of Symbols

A -- set of possible actions

B -- bribery paid by the enterprise to the mid-tier governmental unit

L -- likelihood matrix

M -- Markov matrix

P -- monetary payoff to the principal

Q -- payment scheme offered to the informer

R -- monetary payoff to the agent

S -- set of possible states of nature

V -- verifiability structure

W -- expected payoff to the central government

a -- action

k -- index of the likelihood that collusion between enterprise and mid-tier governmental unit is not detected

m -- probability that an enterprise is productive conditional on the mid-tier governmental unit observing ω_1

n -- probability that the mid-tier governmental unit observes ω_1

s -- state of nature

u -- von-Neumann Morgenstern utility function of the decision maker

y -- output level of an enterprise

Θ -- set of payoff-relevant states of nature

Ω -- signal

γ -- choice mechanism

λ^E, λ^I -- weights of the enterprise's and the mid-tier governmental unit's payoffs, respectively, in the central government's payoff function

ν -- prior probability that an enterprise is productive

π -- probability function defined over the set of possible events

INTRODUCTION

In the economic literature of decision making under uncertainty, we investigate how a decision maker chooses the optimal action to maximize his expected utility when he does not know the state of nature. One of the most fundamental insights in the literature is that the decision maker will be (weakly) better off if he can costlessly observe a signal of the state of nature (Laffont 1989; 56-66). The intuition is that the decision maker can always ignore the signal and achieve at least what he can achieve when there is no signal at all. Another much more important insight in the literature is that if a signal is sufficient[2] in statistical sense for another signal, then the former signal is more informative in economic sense than the latter one in *all* decision making problems. This powerful statement is called the Blackwell theorem (Blackwell, 1951). The Blackwell theorem fills the gap between the statistical concept of sufficiency and the economic concept of informativeness, and hence provides a partial ordering for signals which has economic importance.

However, the practical usefulness of the Blackwell theorem is limited. The Blackwell theorem compares the informativeness of costlessly observed signals. But in most of the economic settings, signals are not costlessly observed. In particular, in a large number of economic situations, the decision maker does not observe the signal by himself directly, but rather buys the signal from a third party instead. Mechandizers buy quality-check reports of certain commodities, employers request recommendation letters from job-seekers, prospective investors read financial statements of corporations that are going to issue new stocks..., etc. These are but a few examples of decision makers obtaining signals about the state of nature from some third parties. The very difference between observing a signal oneself and buying a signal from a third party is that while one cannot cheat himself, the third party can cheat the decision maker. So decision makers seldom treat the signals they obtain in the way the Blackwell theorem implicitly pre-supposes. The Blackwell theorem is applicable only when

2 To be defined in chapter one.

the third party concerned is honest, or just for whatever reason cannot cheat, which is surely a very restrictive, if not too unrealistic, case.

This thesis is an attempt to fill this gap. We argue that when the signals are provided by some third parties called informers, the decision makers' choice over signals is in essence a choice over informers. We will develop another informativeness concept, the v-informativeness, to compare informers, in contrast to that of Blackwell (which we would like to call it the Ω-informativeness) which is used to compare signals. The problem solved by Blackwell can therefore be viewed as a subset of our more general problem, in which the third party concerned will not cheat. Though we have not succeeded in advancing necessary and sufficient conditions for v-informativeness of informers in the most general case, we do have derived some sufficient (and sometimes necessary) conditions in some restricted cases.

To demonstrate how distinguishing the two concepts of informativeness can improve our understanding of real life economic phenomenon, we will also investigate into the Chinese central government's problem of signing responsibility contracts with its state enterprises. The Chinese central government's problem can be read as a problem of choosing informers because the Chinese central government has to rely on its informers -- the local governments -- to determine the terms of contracts. We show by a very simple model how considering the Ω-informativeness only may misguide the central government.

While this thesis can be viewed as a supplement of the Blackwell theorem, it can also be viewed as a contribution to the literature of information cost.

The emphasis on the cost of information put by economists seems to seek its historic roots from at least two sourses. The first source is the debate on the relative superiority of market economy versus central planning. It was argued intuitively by advocates of market economy that the price system employed by the market is more informationally superior in the sense that it economizes the cost of information gathering and subsequent computation. However, before this intuition can be formalized and verified, there remain numerous gaps to be filled (Marschak, 1987). In particular, as market equilibrium is approached only iteratively under the price system, a market economy may not be able to economize the cost of information transmission though it may really economize the cost of information gathering.

Therefore we at least need an explicit measure of the costs of information gathering and transmission if we are to aggregate these two costs in order to verify the intuition of the market economy advocates. This idea had been explicitly voiced by Arrow (1974), "... what was left obscure is a more definite measure of information and its costs, in terms of which it would be possible to assert the superiority of the price system over a centralized alternative... if we are going to take informational economy seriously, we have to add to our usual economic calculations an appropriate measure of the costs of information gathering and transmission." (p.5)

The second root of the literature of information cost is the team theory. Briefly speaking, the problem of a team is to determine which team member is to collect what information (or observe what signals), and to communicate with which other fellow members, and to make decisions according to what decision rule, in order to maximize some collective objective functions of the whole team. Again, an explicit measure of the costs of information gathering and transmission is crucial to this analysis, because the absence of information cost will unrealistically imply that all team members should collect as much information as possible. In fact, due to the immaturity of the literature of information cost, "most of the theory of teams to date has concentrated on choosing optimal decision structures[, that is, the choice of decision rules,] for a given information structure[, that is, the assignment of signals to team members], rather than optimizing on information structures, thereby avoiding explicit consideration of cost functions." (Arrow, 1985:304)

It turns out that the derivation of a direct measurement of the costs of information is extraordinarily difficult. Most of the papers in the literature heavily borrow the existing information theory to construct their functions of information cost, either that of gathering or transmission. According to the information theory, if the set of possible states of nature and the probability distribution of the states of nature are known, then the minimum expected number of letters needed to encode the true state of nature can be calculated. Let $S=\{s_1, s_2, \ldots\}$ denote the countably infinite set of possible states of nature, $P=\{p_1, p_2, \ldots\}$ denote the corresponding probability distribution of the states of nature, and $A=\{a_1, a_2, \ldots, a_k\}$ denote the finite collection of alphabet used to encode the true state of nature. Then the minimum expected number of letters needed to encode the true state of nature can be approximated by the entropy $H(P)$ of P, where:

$$H(P) = -\sum p_i \log_k p_i .$$

This result can be directly borrowed to construct a measure of the cost of information gathering. Suppose one has the following procedure for identifying the true state of nature. He first divided the set of possible states of nature in any desired way into k parts, and it is possible to identify in which part the true state of nature lies. He can then divided the remaining subset of states of nature into k parts, and repeat the above procedure again and again, until the true state of nature is singled out. According to the information theory, the minimum expected number of times the above procedure has to be repeated in order to single out the true state of nature can be approximated by $H(P)$. If the cost of information gathering is proportional to the number of times the above procedure is to be repeated, then $H(P)$ can be used as a proxy of the cost of information gathering. Arrow (1985) suggests one such measure of cost of information gathering with $k = 2$.

Similar way is employed to construct a measure of the cost of information transmission. Suppose the information about the true state of nature is to be transmitted with a language with k available different letters. Then the information theory suggests that the minimum expected number of letters to encode the information can be approximated by $H(P)$. If the cost of information transmission is proportional to the number of letters to be transmitted, the minimum expected cost of transmitting a piece of information will then be a function of $H(P)$. Oniki (1986) applies one so constructed cost function to argue that market economy incurs a lower cost of information transmission than central planning.

Yet not all models model the cost of information transmission by capturing the direct costs (time and resources) incurred. Sah and Stiglitz (1986) argue that there are also indirect costs which "result from the inevitable contamination that occurs in the process of information communication." (p.717) Sah and Stiglitz (1986) investigate the implication of these indirect costs for the choice of the architecture of an economic system or organization. In particular, they argue that if the objective of an organization is to pick good projects from a pool of good and bad projects, a polyarchical form of architecture will admit more good projects as well as bad ones, while a hierarchical form of architecture will do the opposite. So the relative cost of rejecting too many good projects as compared to that of admitting too many bad projects will determine the optimal architecture of the organization.

While Sah and Stiglitz (1986) model the indirect costs of information transmission as exogenously given, Green and Laffont (1986) go a step further to endogenize even these costs. The information theory is again called into service in the due course. Suppose the information about the true state of nature has to be transmitted R_0 times every unit time, and the transmission technology only allows R_1 letters to be transmitted every unit time. Information transmission incurs direct costs in the sense that the installation of the transmission technology is costly, and the installation cost is higher for larger R_1. According to the information theory, the minimum R_1 that guarantees perfect transmission can be approximated by $R_0H(P)$. If $R_1 < R_0H(P)$, some different states of nature may have to be encoded into the same string of letters during the transmission, and some errors will therefore arise in the process of transmission. So the choice of transmission technology involves the trading off of the installation cost against the indirect costs arising from transmission errors.

One of common features of the above models is that costs of information are always related to the amount of information -- it is exactly the information theory which provides a measure for the amount of information is called into service in almost all the above models. Our work in this thesis represents a completely different approach to the information cost problem. In our model, a decision maker will incur a cost when he collects information because additional constraints will be imposed to his set of feasible actions. This deters the decision maker from collecting as much information as possible.

The structure of this thesis is as follows: Chapter one develops the theoretical framework used to compare informers, and derives some preliminary insights on the v-informativeness of informers. Chapter two investigates into the Chinese central government's problem of signing responsibility contracts. Though chapter one and chapter two are related in the sense that chapter two exhibits a working example for some results in chapter one, the two chapters can nevertheless be read independently without loosing continuity.

CHAPTER ONE: ON V-INFORMATIVENESS OF INFORMERS

Section one: Introduction

Consider a decision maker with a von-Neumann Morgenstern utility function $u(a,\theta_i)$ defined over $A \times \Theta$, where A is the set of possible actions and Θ is the set of possible (payoff-relevant) states of nature, $(\theta_1, \theta_2, \ldots, \theta_I)$. Assume that the decision maker cannot observe the state of nature. So the problem of an expected-utility-maximizing decision maker is:

$$\max_{a \in A} \sum_{\theta_i \in \Theta} \pi(\theta_i) u(a, \theta_i),$$

where $\pi(\theta_i)$ is the probability that the state of nature θ_i has realized. ($\sum_{\theta_i \in \Theta} \pi(\theta_i) = 1$; $\pi(\theta_i) > 0$, $\forall \theta_i \in \Theta$.)

Suppose now the decision maker can costlessly observe a signal before acting. Denote the signal by Ω, which is in fact a set of possible values observed by the decision maker, $(\omega_1, \omega_2, \ldots, \omega_J)$. The decision maker can then improve his choice of action by making his choice contingent on the value of the signal he happens to observe. Formally, the problem of the decision maker becomes:

$$\max_{a(\omega_j)} \sum_{\theta_i} \pi(\theta_i) \sum_{\omega_j} \pi(\omega_j | \theta_i) u\big(a(\omega_j), \theta_i\big),$$

where $\pi(\omega_j | \theta_i)$ is the probability that value ω_j will be observed conditional on that θ_i has realized.

Since the decision maker can always choose $a(\omega_j) = a^*$, where a* is the solution to the decision-making-without-signal problem, the decision maker can never be worse off with signal Ω. This is one of the fundamental insights in the literature of decision making under uncertainty.

Suppose now the decision maker can choose between two signals. Can we tell which one should he choose? Since a signal will have different values for different decision makers, and a signal that is valuable for some decision makers may have little value to other decision makers, it is therefore impossible to introduce a total ordering for the signals according to their values to decision makers. Nevertheless, Blackwell suggests that there does exist a partial ordering of signals that can be economically meaningful. The now well known

Blackwell theorem states that the partial ordering of signals based on the statistical concept of sufficiency is equivalent to an economic partial ordering of signals based on the values of the signals to the decision makers.

Let $\wp(\Omega;A,u)$ denote the maximized expected utility of the decision maker with utility function u and action set A when he makes use of the signal Ω. Ω_1 is said to be more informative than Ω_2 if $\wp(\Omega_1;A,u)$ has values higher than or equal to that of $\wp(\Omega_2;A,u)$ for all decision makers. In the subsequent discussion another concept of "informativeness" will be defined which is called v-informativeness. To clearly distinguish the "informativeness" used here from the one to be defined subsequently, we will called the one used here as Ω-informativeness. It is sufficient by now to remember that Ω-informativeness is a partial ordering of signals (while v-informativeness is a partial ordering of informers).

Formally we have:

Definition Ω_1 is *more Ω-informative than* Ω_2 if $\wp(\Omega_1;A,u) \geq \wp(\Omega_2;A,u)$ for all A and u.

Statistically, we say that Ω_1 is sufficient for Ω_2 if Ω_2 is generated simply by "adding noise" to Ω_1. Let $\mathbf{L}(\Omega)$ denote the likelihood matrix of signal Ω, that is:

$$\mathbf{L}(\Omega) = \left(\pi(\omega_j|\theta_i)\right)_{I\times J},$$

where i is the column number and j is the row number. Then we have:

Definition Ω_1 is *sufficient* for Ω_2 if there exists a Markov matrix,[3] $\mathbf{M}$, such that:

$$\mathbf{L}(\Omega_2) = \mathbf{L}(\Omega_1)\mathbf{M}.$$

To see how this definition conveys the idea of "adding noise," we can interpret the component of in the p-th row and q-th column of $\mathbf{M}$ as $\pi(\hat{\omega}_q|\omega_p)$, where ω_p is the p-th value of Ω_1 and $\hat{\omega}_q$ is the q-th value of Ω_2. If $\mathbf{M}$ exists, the conditional probabilities, $\pi(\hat{\omega}_q|\omega_p)$'s, will then be constants independent of the state of nature. So signal Ω_2 can be interpreted as signal Ω_1 adding a "noise" independent of the state of nature.

The Blackwell theorem then states that the economic partial ordering and the statistical partial ordering of signals are in fact equivalent.

---

3 A matrix $\mathbf{M}$ is a Markov matrix if $\mathbf{M} \geq \mathbf{0}$ and $\mathbf{Me} = \mathbf{e}$, where $\mathbf{e}$ is a column vector each component of which is 1.

Blackwell Theorem[4] Ω_1 is more Ω-informative than Ω_2 if and only if Ω_1 is sufficient for Ω_2.

Following Blackwell's spirit, Arrow (1992) goes a step further to develope the informational equivalence of signals. The definition of informational equivalence comes naturally from Blackwell's Ω-informativeness:

Definition Ω_1 is *informationally equivalent* to Ω_2 if Ω_1 is more Ω-informative than Ω_2 and Ω_2 is more Ω-informative than Ω_1.

Arrow (1992) argues that two signals are informationally equivalent if and only if one of the following two cases holds. The first is the signals differ from each other only in the names assigned to the signal realizations. The second is that if both signals abandon all their redundant realizations (for example, those realizations that will never realize), they then differ from each other only in the names assigned to the signal realizations. While the *if* part of this result is obvious, the *only if* part is both strong and non-trivial. Arrow's (1992) result can be formally summarized by the following two definitions and two theorems:

Definition A matrix is *irredundent* if no columns are zero and no two columns are collinear.

Definition A signal, S, is *s-minimal* if and only if **L**(S) is irredundant.

Theorem 2.1 For any signal, S, there exist a s-minimal signal, S', such that S is informationally equivalent to S'.

Theorem 2.2 Two s-minimal signals are informationally equivalent if and only if they are the same apart from a relabeling of the signal realizations.

Suppose now the decision maker cannot observe the signals directly, but has to rely on some informers to inform him about the realized values of the signals instead. Can we tell which signal should he choose now? Is the Blackwell theorem still useful in this situation?

The answer to the last question must be negative, because it is apparent that comparing the signals alone must be far from enough now. Even when two informers can observe the same signal, they may have very different ability to verify their reports to the

4 For a simple proof see Cremer (1982).

decision maker and hence may have very different degree of "credibility." Therefore, rather than choosing a more Ω-informative signal, the decision maker instead chooses a more "informative" informer, with the signal being one of the characteristics of the informer. To emphasize the difference between the "informativeness" of a signal and that of an informer. We use the term v-informativeness when we are referring to informers

Our task in this chapter is to derive some preliminary insights on comparing the v-informativeness of informers. Section two will introduce the concept of verifiability structure which will be the corner-stone of our comparison. Section three introduce the formal structure of the class of principal-informer-agent problems which includes most of the economic settings in which the signals the decision makers consult is provided by some informers. Section four presents two theorems on the v-informativeness of informers in the framework of principal-informer-agent problems.

Section two: Verifiability structure

Consider two extreme types of informers. The first type of informers can verify all possible reports they make to the decision makers. For example, suppose the weather of a distant city is of concern to the decision maker. If an informer can obtain the weather reports of the observatory there, then he not only can provide the decision maker with a valuable signal of the weather there, but can readily verify whatever he says by simply showing the decision maker the weather reports he obtains as well. In this case, employing the informer is no different from directly observing the signal, because the decision maker does not need to worry that the informer will cheat for whatever reason. Whenever the informer makes a report, the decision maker can at the same time ask for verification, and only accept the report if the informer can verify it. Since the decision maker knows that the informer can verify all reports provided that the report is true, he can therefore deduce that the informer must be lying if the informer fails to verify his report. Knowing this logic, the informer will not dare to lie. The decision maker's choice over informers hence becomes trivial. By the Blackwell theorem, the decision maker should never choose informers whose signals are less Ω-informative than the others[1].

Then consider another extreme case. The second type of informers can verify none of the values of the signals they observe to the decision makers. For example, the weather observer may not be able to verify what he claims he has observed simply because he is not observing with the help of any weather report. In this case the decision maker can never know whether the informer he employs is telling the truth or just cheating. Even if the informer has no incentive to cheat, he does not have any incentive to tell the truth either. So the decision maker's best strategy in such situation is to ignore any report made by the informer and act as if there is no informer at all. Here the decision maker's choice over informers is trivial again. The Blackwell theorem is totally useless. No matter how Ω-informative the signal observed by an informer is, the decision maker should never employ him.

Yet there are numerous cases in between the two extreme cases above. Most of the economic settings are characterized by a situation in which the informers can prove only some of the values of the signals they observe and not the others. Take the above weather-informer as an example again. Suppose the observatory he relies on will collapse and all scientists inside will die if there is an earthquake at the city. Then the claim of the weather-informer that an earthquake has occurred becomes non-verifiable. When the informer claims that there has been an earthquake, the decision maker simply do not know whether the informer is hiding the weather reports and lies, or there is really an earthquake.

The fact that some informers cannot prove all of their possible claims can greatly complicate the decision maker's choice over informers because two informers may have very different values to a decision maker even when the signals they observe are the same, provided that they have different subsets of values of their commonly-observed signal non-verifiable. For example, a weather-informer who cannot verify a claim of earthquake may have very different value to a decision maker from one who cannot verify a claim of shower. We can say that these two informers are different not in the signals they observe, but in the *verifiability structures* of their signals.

Therefore, in addition to the signal, the verifiability structure is another major characteristic of an informer that a decision maker should consider in his choice over informers. The first step to compare the v-informativeness of informers is hence to formalize the above idea of verifiability structure.

To formalize the concept of verifiability structure, we have to first formalize that of "being able to verify." It seems natural to define "to verify" as the presentation of evidence(s) which can prove that a certain event has happened, and to define the verifiability structure of a signal as a partition of the set of all possible values of that signal into two subsets, one for all verifiable values and the other for the non-verifiable ones. However, while this may seem enough to describe all possible verifiability structures of a weather-informer in the above example, it may be far from adequate in many other cases. In particular, it may happen that some values of a signal are neither perfectly verifiable nor totally non-verifiable.

For example, the number of unpublished creative papers an Ph.D. student has written may well be a signal of how productive he is. Being an informer of any may-be-employer of that student, the student's supervisor can partially verify his claim that the student has written a certain number of papers by showing exactly that number of papers authored by that student. However, the verification is by no means perfect, because the supervisor cannot prove that the number of papers written is no more than that he claims; or in other words, he can never prove that he has not hidden any paper of his student. In this case, the report of no paper is more "non-verifiable" than that of one paper, which is in turn more "non-verifiable" than that of two papers..., etc.

To capture this idea of "relative verifiability," we suggest the following definition of "verifiable." We says that an informer can *verify* a particular value of a signal if for every other values (he claims to have) not realized, he can advance evidences to prove that it has really not realized.[5,6] The corresponding definition of verifiability structure is therefore a class of subsets of a certain signal.

---

5 I am indebted to Wing-Chung Pun for suggesting this verify-by-piecewise-denying definition, which eliminate a fatal defect of my eariler version.

6 It is certainly not the only possible technology of verification. In fact we can easily raise a lot more different technologies of verification, each of which has interesting features not yet captured by our discussion. We use this definition only because it is a good starting point for this study in the sense that it allows for imperfect verifiability and yet will not add too much complication to our subsequent discussion.

How the definitions suggested above can capture the idea of "relative verifiability" can best be illustrated with the help of figures 1 to 4. For a four-value signal, that is, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, the following four figures represent four of the many possible verifiability structures. In the figures, each black spot represents one value of the signals, ω_j. Each circle therefore encloses a verifiable subset of values. The whole family of circles thus makes up a verifiability structure, for example, in figure 3, the verifiability structure consists of the following elements: ϕ, $\{\omega_2\}$, $\{\omega_2, \omega_3\}$, $\{\omega_2, \omega_3, \omega_4\}$, and Ω. The interpretation of any encircled subset is that the informer concerned is able to advance separate evidences to prove that each of the values outside the encircled subset has not realized, provided that the observed value (the value that has actually realized) is in the encircled subset.

Figure 1 illustrates one of the most trivial verifiability structures. Here none of the values is to any extent verifiable. As we have discussed before, informer with such a verifiability structure is of no value to any decision maker.

Figure 2 represents another trivial case that all values are perfectly verifiable. It resembles the first extreme type of informers discussed above. Yet figure 2 is far from a full exposition of the verifiability structure concerned. Consider a case in which the informer observes value ω_2. Surely he cannot pretend that he has observed value ω_1, as there is a circle enclosing spot ω_1, indicating that any one who really observes value ω_1 can readily prove it. Similarly the informer cannot pretend that he has observed value ω_3. These jointly mean that the informer cannot pretend that he has observed "values ω_1 or ω_3," and therefore there should be a circle enclosing at least spots ω_1 and ω_3 but excluding spot ω_2. In fact, *in a complete set-diagrammatic exposition of verifiability structure, union of any two encircled subsets should also be circled.* So figure 1 is a complete exposition while figure 2 is not.

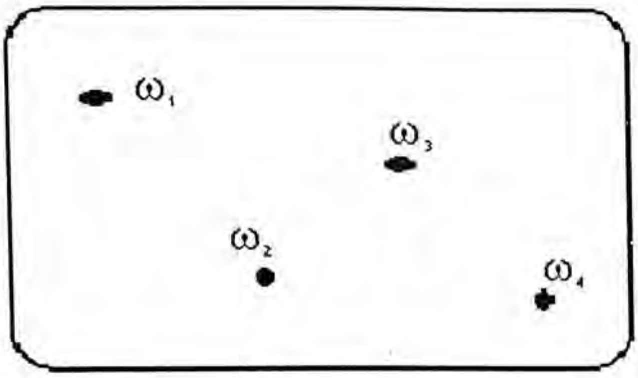

Fig. 1

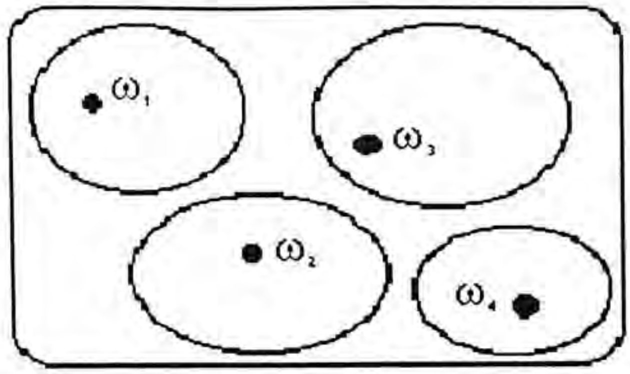

Fig. 2

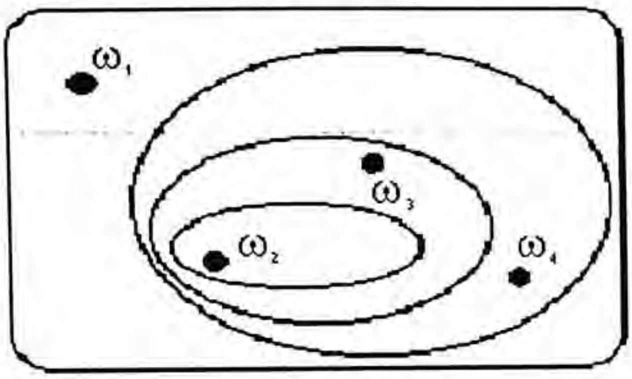

Fig. 3

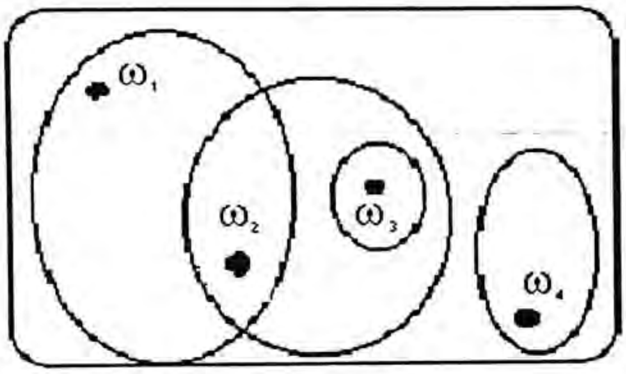

Fig. 4

Figure 3 represents a situation very much resembles the example of the Ph.D. student. Here value ω_2 is perfectly verifiable, but not values ω_1, ω_3 and ω_4. For example, when the informer observes value ω_3, he can advance evidence that he has not observed values ω_1, and evidence that he has not observed value ω_4, but not that he has not observed value ω_2. In other words, when he claims that he has observed value ω_3, we do not know whether he has actually observed value ω_2 but pretend that he has observed value ω_3 instead.

Note that though there is a circle enclosing values ω_2, ω_3 and ω_4, there is no circle enclosing value ω_1. Actually *we do not require that the complement of a verifiable subset is also verifiable*. This is especially clear in the weather-informer example: Though the weather-informer can verify "no earthquake," he cannot verify its complement, that is "there is earthquake."

Figure 4 exemplifies another more complicated situation. At first sight it seems that an informer observing value ω_3 can pretend that he has observed value ω_2. But it is not the case. In fact, an observer who has really observed value ω_2 can readily prove it by first proving that he has observed "values ω_1 or ω_2," and then proving that he has observed "values ω_2 or ω_3." In other words, figure 4 is also incomplete in the sense that there should at least be one more circle enclosing spot ω_2. As a rule, *in a complete set-diagrammatic exposition of verifiability structure, intersection of any two encircled subsets should also be circled.*

We can now summarize our discussion on verifiability structure in the following formal definition.

Definition A *verifiability structure*, $\mathbf{V}(\Omega)$, of a signal Ω is a family of subsets of Ω, such that:

(1) $\phi, \Omega \in \mathbf{V}(\Omega)$; and

(2) if $V_1, V_2 \in \mathbf{V}(\Omega)$, then $V_1 \cap V_2 \in \mathbf{V}(\Omega)$, and $V_1 \cup V_2 \in \mathbf{V}(\Omega)$.

Note that (2) restates that intersection and union of any two verifiable subsets are also verifiable. (1) is here just for completeness; the role of (1) is to guarantee that **V** is closed under intersection and union.

Section three: Principal-informer-agent problem

In this section we will formally introduce the family of principal-informer-agent problems. This will become the framework for our comparison of informers. The family of problems introduced here differs from that introduced in section one in two ways. The first is of course that we have to formally introduce an informer who is to report the signal to the principal into our problems. The second is that we introduce an agent who will be affected by the action of the principal, and hence may want to bribe the informer to mis-report the signal.

An informer is characterized by two things. The first is the signal he observes; informers observing different signals are undoubtedly different in the eyes of the decision makers. The second is the verifiability structure of the signal the informer observes; difference in verifiability structures makes informers differ from each other even though they may observe the same signal.

Yet if the informer has no conflict of interest with the decision maker, the informer has no incentive to cheat the decision maker. The reason why we do care about whether the informer will cheat is that sometimes the interest of the decision maker and the informer may not coincide. We can of course introduce an utility function to the informer so as to make the informer care about the decision maker's action. But this adds unnecessary complexity to the comparison of informers, as we will have to regard informers with different utility functions as different, though they may observe the same signal and have the same verifiability structure. Therefore we choose to introduce an agent affected by the decision maker's action to the problems instead. When the agent is adversely affected by the decision maker's action, he may contemplate bribing the informer so that the informer will lie and misguide the decision maker's action in favor of the agent. If the informer maximizes the bribery he receives, he will then cheat the decision maker. This captures the essence of many real life economic settings in which decision makers hesitate to believe what the informers say with the worry that the informers may report in some other parties' interest. To model the way possible collusion between the informer and the agent affects the principal, we follow the pioneering work of Tirole (1986).

This way of modeling also has the advantage that the family of problems can then include many popular problems in the economic literature as special cases; examples of them include decision making problems under uncertainty (that family of problems on which Blackwell derives his theorem), the principal-agent problems with hidden information, and Tirole's (1986) principal-supervisor-agent problem. We will first introduce the formal structure of the family of principal-informer-agent problems, and then show by some examples how it includes the above mentioned problems as special cases.

Let S denote the set of possible states of nature, with s as its typical element. We assume that S is finite. $\pi(.)$ is the probability distribution function defined over 2^S,[7] with $\sum_{s \in S} \pi(s) = 1$ and $\pi(s) > 0$ for all $s \in S$.

Definition A *payoff-relevant state of nature*, denoted by θ_i, is a partition set of S. Denote by Θ the set of all payoff-relevant states of nature, that is $\Theta = \{\theta_i\}$.

Consider a principal who is to choose an action, a, from the set of possible actions, A. Having taken action a, the principal can receive a monetary payoff amounts to $P(a, \theta_i)$. We say that the vector (A, P) fully describes a principal.

7 2^S denotes the power set of S, that is, $2^S := \{X | X \subseteq S\}$.

Definition A *principal* is a vector (A, P), where P is a payoff function mapping from the action set and the set of payoff-relevant states of nature to the real line: $P: A \times \Theta \to \Re$.

We assume that the principal is risk-neutral and maximizes his expected *net* payoff (his monetary payoff net of any possible payment). We also assume that the principal cannot observe the state of nature before he chooses the action. So his problem is to:

$$\max_{a \in A} \sum_{\theta_i} \pi(\theta_i) P(a, \theta_i).$$

An agent is one who is affected by the action taken by the principal and knows the (payoff-relevant) state of nature.[8] We assume that an agent cannot verify any report about the (payoff-relevant) state of nature to the principal, so that he can never become an informer which is to be defined below. Given action a and state of nature θ_i, the agent will have a monetary payoff amounts to $R(a, \theta_i)$. We say that the function R fully describes an agent.

Definition An *agent* is a payoff function, denoted by R, mapping from the set of payoff-relevant states of nature and the action set to the real line: $R: \Theta \times A \to \Re$.

We assume that the agent maximizes his monetary payoff. For most parts of this chapter we do not need to make any assumption on the agent's risk attitude. However, when we go to theorem 1.2 in section four of this chapter, we will assume that the agent is risk-neutral since this lends us much convenience.

Before choosing the action, the principal can consult an informer, who can costlessly observe a signal of the state of nature.

Definition A *signal* of Θ is another partition of S, denoted by Ω, with ω_j its typical element, i.e. $\Omega = \{\omega_j\}$.

Note that Ω can be equal to or not equal to Θ.

We say that the vector $(\Omega, V(\Omega))$ fully describes an informer.

Definition An *informer* is a vector $(\Omega, V(\Omega))$.

We assume that the informer maximizes his monetary income. Similar to the case of the agent, we do not need to make any assumption on the informer's risk attitude in most parts

[8] This definition is clearly very different from the conventional one. However, the two are in fact compatible. See the examples below.

of this chapter. But it will lend us much convenience by assuming that the informer is risk-neutral when we go to theorem 1.2 in section four of this chapter.

Difference in verifiability structures enables different informers observing the same value of the same signal to have different choices of reports.

Definition A report, denoted by μ, is *feasible* if $\mu \in \Omega$. Clearly Ω is then also the set of feasible reports.

This definition is saying that the informers are supposed to report what they can observe, rather than other things else.

It is assumed that $\mathbf{V}(\Omega)$ is common knowledge,[9] in particular $\mathbf{V}(\Omega)$ is known to the decision maker. Hence, not all feasible reports will be accepted by the decision maker. If some particular values of a signal are known to be verifiable provided they have realized, then a report that one of these values has realized will not be accepted by the decision maker if the report is not accompanied by the expected verification. This idea is captured by the following definition.

Definition Given ω_j and $\mathbf{V}(\Omega)$, a feasible report, μ, is *inadmissible* if there exists $V \in \mathbf{V}(\Omega)$ such that $\omega_j \notin V$ and $\mu \in V$.

If such an V exists, then the decision maker can ask the informer for verification that ω_j has not occurred, which is impossible to exist given ω_j. Hence, μ is inadmissible.

Definition Given ω_j and $\mathbf{V}(\Omega)$, a feasible report, μ, is *admissible* if it is not inadmissible.

Denote by $M(\omega_j)$ the set of all admissible reports when ω_j has realized. Note that: (1) $M(\omega_j) = \{\mu \in \Omega | \text{if } \mu \in V \in \mathbf{V}(\Omega), \text{ then } \omega_j \in V\}$. In other words, when ω_j has realized, μ is admissible if there is no verification that can distinguish μ and ω_j. (2) $\omega_j \in M(\omega_j)$, for all $\omega_j \in \Omega$.

Example In figure 1, given ω_3 is the observed value, ω_2 is not an admissible report because $\{\omega_2\}$ is in $\mathbf{V}(\Omega)$ but $\omega_3 \notin \{\omega_2\}$. However, ω_1 is an admissible report because the only element of $\mathbf{V}(\Omega)$ that contain ω_1 is Ω, but Ω also contains ω_3.

---

9 Here, "common knowledge" is used in a non-technical way, namely, everyone knows $\mathbf{V}(\Omega)$, everyone knows that everyone knows $\mathbf{V}(\Omega)$, and so on, *ad infinitum*.

Example In figure 1, $M(\omega_3) = \{\omega_1, \omega_3, \omega_4\}$.

A payment scheme offered by the principal to the informer consists of an agreement over what payment be paid to the informer conditional on the report submitted. Let $Q(\omega_j)$ denote the payment scheme. A natural restriction on the set of possible Q is that $Q \geq 0$ for all $\omega_j \in \Omega$.

We assume that the agent knows what the informer has observed and may try to bribe the informer to make a false report to the principal which is more favorable to the agent.

The principal's problem becomes to cho[illegible] payment scheme and a choice mechanism to maximize his expected monetary payoff net of payment paid to the informer, taking into account the possible collusion between the informer and the agent.

Definition A *choice mechanism*, $\gamma(\mu)$, is a function mapping from the set of feasible reports to the action set.

We say that the vector (A, P, Ω, V, R) fully describes an principal-informer-agent problem.

Definition An *principal-informer-agent problem* is the vector (A, P, Ω, V, R) in which the principal is to solve the problem:

$$\max_{Q(\cdot),\gamma(\cdot)} \sum_{\omega_j} \pi(\omega_j) \sum_{\theta_i} \pi(\theta_i|\omega_j)\left[P(\gamma(\omega_j), \theta_i) - Q(\omega_j)\right]$$

subject to:

(IIC's) $\quad Q(\omega_j) \geq Q(\mu) + \Delta_{j\mu}, \ \forall \mu \in M(\omega_j), \ \forall \omega_j \in \Omega,$

where $\quad \Delta_{j\mu} = \max\left\{\max_{\theta_i}\left\{R(\gamma(\mu), \theta_i) - R(\gamma(\omega_j), \theta_i)\right\}, 0\right\}.$

Here, $\Delta_{j\mu}$ is the maximum bribery the agent is willing to pay the informer in order to induce the informer to report μ when actually the informer has observed ω_j. Constraints (IIC's) hence represents the informer's incentive compatibility constraints.

The following examples illustrate that the class of principal-informer-agent problems consists of many decision making problems widely studied in the economic literature.

Example It is clear that the principal-informer-agent problem is a direct generalization of Tirole's (1986) principal-supervisor-agent problem. A principal-supervisor-agent problem is a principal-informer-agent problem with the principal's action set equals to the set of possible contracts between the principal and the agent, constrained by that the contracts satisfy the agent's individual rationality constraint and incentive compatibility constraints. Moreover, the supervisor in the principal-supervisor-agent problem is an informer with the following specific signal and verifiability structure:

$$\Omega = \{\omega_0, \omega_1, \omega_2, \ldots, \omega_n\},$$

$$\pi(\theta_i|\omega_j)\begin{cases} = 1 \text{ if } i = j \\ = 0 \text{ if } i \neq j, j \neq 0, \\ > 0 \text{ if } j = 0 \end{cases}$$

$$\mathbf{V}(\Omega) = \{\phi, 2^{\Omega\backslash\omega_0}, \Omega\},$$

where $\Omega \backslash \omega_0 = \{\omega_1, \omega_2, \ldots, \omega_n\}$. The intuitive interpretation of ω_0 is that when the informer observes ω_0, he observes nothing and can prove or disprove nothing. That is, ω_0 represents "no information."

Example When $\mathbf{V} = 2^{\Omega}$, we have $M(\omega_j) = \{\omega_j\}$ for all $\omega_j \in \Omega$. When ω_j has realized, the informer can only report ω_j. In other words, the informer cannot cheat. Then, (IIC's) are trivially satisfied. The principal-informer-agent problem becomes analytically equivalent to a decision making problem under uncertainty with the help of the signal Ω. What the decision maker needs to do to motivate the informer to truly report the observed signal is to pay the informer a minimal amount above the latter's reservation wage.

Example The general principal-agent problem with hidden information is a special case of the principal-informer-agent problem in which the informer has the verifiability structure, $\mathbf{V} = \{\phi, \Omega\}$, and the action set is the set of possible contracts between the principal and the agent, constrained by that the contracts satisfy the agent's individual rationality constraint and incentive compatibility constraints. Note that in this case $M(\omega_j) = \Omega$ for all $\omega_i \in \Omega$, which means that the informer can report anything he likes no matter what he actually has observed and he cannot verify any value of signal he claims to have observed. The informer is worthless. Therefore, we neglect this informer in the principal-agent literature.

The formal structure of principal-informer-agent problems provides the framework for us to compare the v-informativeness of informers. Yet we have to first define formally what do we exactly mean by v-informativeness.

Let $\wp(\Omega, V; A, P, R)$ denote the maximized payoff to the principal in the principal-informer-agent problem (A, P, Ω, V, R). We can now define the v-informativeness of an informer.

Definition (Ω_1, V_1) is *at least as v-informative as* (Ω_2, V_2) if for all A, P and R, $\wp(\Omega_1, V_1; A, P, R) \geq \wp(\Omega_2, V_2; A, P, R)$.

It can be easily seen that "at least as v-informative as" is a partial ordering. That is, (Ω_1, V_1) and (Ω_2, V_2) may not be comparable according to the relation.

Definition (Ω_1, V_1) is *more v-informative than* (Ω_2, V_2) if (Ω_1, V_1) is at least as informative as (Ω_2, V_2) while (Ω_2, V_2) is not at least as informative as (Ω_1, V_1).

Section four: On v-informativeness of informers

Our task in this section is to derive some preliminary insights in comparing the v-informativeness of the informers. Proposition 1.1 merely re-states the fact that the Blackwell theorem is applicable only when the verifiability structures concerned belong to a very special type -- one that equals to the power set of the signal. In one of the examples above, we have already argued that an informer with such a verifiability structure cannot cheat. Hence the Blackwell theorem can applied.

Proposition 1.1 If $V_1 = 2^{\Omega_1}$ and $V_2 = 2^{\Omega_2}$, then (Ω_1, V_1) is at least as v-informative as (Ω_2, V_2) if and only if Ω_1 is more Ω-informative than Ω_2.

Proof Directly from the Blackwell theorem. **Q.E.D.**

Theorem 1.1 below is the first of two important theorems to be presented in this chapter. Theorem 1.1 will deal with informers observing the same signal, while the assumption that the signals are the same will be relaxed in theorem 1.2.

Theorem 1.1 If $\Omega_1 = \Omega_2$, then (Ω_1, V_1) is at least as v-informative as (Ω_2, V_2) if and only if $V_1 \supseteq V_2$.

Proof (*if* part) $\mathbf{V}_1 \supseteq \mathbf{V}_2$ implies that for all $\omega_j \in \Omega_1 = \Omega_2$, $M_1(\omega_j) \subseteq M_2(\omega_j)$.[10] So for all A, P and R, (IIC's) in $(A, P, \Omega_1, \mathbf{V}_1, R)$ is a subset of that in $(A, P, \Omega_2, \mathbf{V}_2, R)$, and therefore $\wp(\Omega_1, \mathbf{V}_1; A, P, R) \geq \wp(\Omega_2, \mathbf{V}_2; A, P, R)$.

(*only if* part) Suppose $\mathbf{V}_1 \not\supseteq \mathbf{V}_2$. That is, there exists V such that $V \in \mathbf{V}_2$ but $V \notin \mathbf{V}_1$. We first prove that there then exists $\omega_j, \mu \in \Omega_1 = \Omega_2$ such that $\omega_j \notin V$, $\mu \in V$ and $\mu \in M_1(\omega_j)$.

Suppose we cannot find such ω_j and μ. That is for all $\omega_j \notin V$ and $\mu \in V$, there exist an element in $\mathbf{V}_1$, denoted by $X(\omega_j, \mu)$, such that $\omega_j \notin X(\omega_j, \mu)$ and $\mu \in X(\omega_j, \mu)$. For all $\omega_j \notin V$, define the subset $Y(\omega_j)$ by $Y(\omega_j) := \bigcup_{\mu \in V} X(\omega_j, \mu)$. Note that $Y(\omega_j)$ is a superset of V, because for all $\mu \in V$, there exists $X(\omega_j, \mu)$ such that $\mu \in X(\omega_j, \mu) \subseteq Y(\omega_j)$. Note also that $Y(\omega_j)$ does not contain ω_j, because for all $\mu \in V$, $X(\omega_j, \mu)$ does not contain ω_j. Define another subset Z by $Z := \bigcap_{\omega_j \notin V} Y(\omega_j)$. Note that Z is also a superset of V, because for all $\omega_j \notin V$, $Y(\omega_j)$ is a superset of V. Note also that for all $\omega_j \notin V$, $\omega_j \notin Z$ because there exists a $Y(\omega_j)$ which does not contain ω_j. Since Z is a superset of V, and does not contain any $\omega_j \notin V$, we conclude that $Z = V$. Since for all $\omega_j \notin V$ and $\mu \in V$, $X(\omega_j, \mu)$ is an element of $\mathbf{V}_1$, and since $\mathbf{V}_1$ is closed under intersections and unions, we have $Z = V \in \mathbf{V}_1$. This contradicts our assumption that $V \notin \mathbf{V}_1$. So we conclude that there exists $\omega_j, \mu \in \Omega_1 = \Omega_2$ such that $\omega_j \notin V$, $\mu \in V$ and $\mu \in M_1(\omega_j)$.

Note that $\omega_j \notin V$, $\mu \in V$ and $\mu \in M_1(\omega_j)$ jointly imply that $\mu \in M_1(\omega_j)$ and $\mu \notin M_2(\omega_j)$.

If we can find some (A,P,R') such that $\wp(\Omega_1, \mathbf{V}_1; A, P, R') < \wp(\Omega_2, \mathbf{V}_2; A, P, R')$, then $(\Omega_1, \mathbf{V}_1)$ is not at least as v-informative as $(\Omega_2, \mathbf{V}_2)$. We are now to find such (A,P,R').

Consider first a special class of problems, $(A, P, \Omega, 2^\Omega, R)$, where $\Omega = \Omega_1 = \Omega_2$ and $\mathbf{V} = 2^\Omega$, such that the problem has unique solution. Let (Q^*, γ^*) be that unique solution to this problem. Since the informer can verify all values he observes, the principal should not worry that he may take bribery from the agent and then mis-report the signal. Moreover, the principal can choose the choice mechanism as if the agent is not affected. So (Q^*, γ^*) must be

10 When an informer can verify more, he has smaller room to cheat.

characterized by $Q^* \equiv 0$ and γ^* independent of R. We assume that γ^* is also characterized by $\gamma^*(\omega_k) \neq \gamma^*(\omega_t)$ for all $k \neq t$.[11] That is the principal will choose different action when the informer reports different reports.

In particular, (Q^*,γ^*) must also solve the problem $(A,P,\Omega,2^\Omega,R')$ where:

$$R'(a,\theta_i) = \begin{cases} 1 \text{ if } a = \gamma^*(\omega_k) \text{ for some } \omega_k \in V \\ 0 \text{ if otherwise} \end{cases}, \text{ for all } \theta_i \in \Theta.$$

We now prove that (Q^*,γ^*) also solves the problem (A,P,Ω,V_2,R'). Since the set of (IIC's) constraints in problem $(A,P,\Omega,2^\Omega,R')$ is a subset of that in problem (A,P,Ω,V_2,R'), it suffices to prove that when the principal applies (Q^*,γ^*) in problem (A,P,Ω,V_2,R'), all (IIC's) constraints are automatically satisfied. Consider first the (IIC's) constraints with $\omega_k \in V$. Since for all $\eta \in M_2(\omega_k)$, $R'(\gamma^*(\eta),\theta_i) \leq 1 = R'(\gamma^*(\omega_k),\theta_i)$ for all $\theta_i \in \Theta$ (because $R'(\gamma^*(\eta),\theta_i)$ is either 0 or 1), so we have $\Delta_{k\eta} = 0$, and $Q^*(\omega_k) \geq Q^*(\eta) + \Delta_{k\eta}$ is therefore fulfilled. Then consider the (IIC's) constraints with $\omega_k \notin V$. Since $M_2(\omega_k) \cap V = \phi$, there does not exist $\eta \in M(\omega_k)$ such that $R'(\gamma^*(\eta),\theta_i) > 0 = R'(\gamma^*(\omega_k),\theta_i)$. Again, we have $\Delta_{k\eta} = 0$, and $Q^*(\omega_k) \geq Q^*(\eta) + \Delta_{k\eta}$ for all $\eta \in M(\omega_k)$ is therefore fulfilled.

Therefore we have $\wp(\Omega,V_2;A,P,R') = \wp(\Omega,2^\Omega;A,P,R')$.

We then prove that (Q^*,γ^*) does not solves the problem (A,P,Ω,V_1,R'). It suffices to prove that (Q^*,γ^*) violates the (IIC's) constraints with ω_j. Since $\omega_j \notin V$, we have $R'(\gamma^*(\omega_j),\theta_i) = 0$. However, $\mu \in V$ and therefore $R'(\gamma^*(\mu),\theta_i) = 1$. So we have $\Delta_{j\mu} = 1$, and $Q^*(\omega_j) = 0 < 1 = Q^*(\mu) + \Delta_{j\mu}$. Therefore at least one (IIC's) constraint is violated.

Since (Q^*,γ^*) is the unique solution to $(A,P,\Omega,2^\Omega,R')$, therefore we have $\wp(\Omega,V_1;A,P,R') < \wp(\Omega,2^\Omega;A,P,R')$. Combining with the result that $\wp(\Omega,V_2;A,P,R') = \wp(\Omega,2^\Omega;A,P,R')$ derived previously, we therefore have $\wp(\Omega,V_1;A,P,R') < \wp(\Omega,V_2;A,P,R')$. **Q.E.D.**

---

11 This assumption can in fact be easily satisfied by construction. Suppose in the originally constructed problem, we have some k and l such that $\gamma^*(\omega_k) = \gamma^*(\omega_t)$. Then we can always construct another problem by adding one more element, $\tilde{a}$, to the original action set such that $P(\tilde{a},\theta_i) = P(\gamma^*(\omega_k),\theta_i)$ for all $\theta_i \in \Theta$. By reassigning $\gamma^*(\omega_k) = \tilde{a}$, we can therefore have $\gamma^*(\omega_k) \neq \gamma^*(\omega_t)$ in the newly constructed problem.

Theorem 1.1 suggests that if two informers can observe the same signal, then the one that can verify all those reports that another informer can verify has a (weakly) higher value to all decision makers.

A direct corollary following theorem 1.1 is:

Corollary 1.1 If $\Omega_1 = \Omega_2$, then $(\Omega_1, \mathbf{V}_1)$ is more v-informative than $(\Omega_2, \mathbf{V}_2)$ if and only if $\mathbf{V}_1 \supset \mathbf{V}_2$.

Proof By definition, $(\Omega_1, \mathbf{V}_1)$ is more v-informative than $(\Omega_2, \mathbf{V}_2)$ is equivalent to that $(\Omega_1, \mathbf{V}_1)$ is at least as v-informative as $(\Omega_2, \mathbf{V}_2)$ and $(\Omega_2, \mathbf{V}_2)$ is not at least as v-informative as $(\Omega_1, \mathbf{V}_1)$, which is true if and only if $\mathbf{V}_1 \supseteq \mathbf{V}_2$ and $\mathbf{V}_2 \not\supseteq \mathbf{V}_1$, which in turn is equivalent to that $\mathbf{V}_1 \supset \mathbf{V}_2$. **Q.E.D.**

Proposition 1.1 and theorem 1.1 jointly have the following corollary.

Corollary 1.2 If $V_1 = 2^{\Omega_1}$, then (Ω_1, V_1) is at least as v-informative as (Ω_2, V_2) if Ω_1 is more Ω-informative than Ω_2.

Proof Directly from proposition 1.1 and theorem 1.1. **Q.E.D.**

Theorem 1.1 can be viewed as a complement to the Blackwell theorem. While the Blackwell theorem restricts our attention to a specific type of verifiability structure, theorem 1.1 restricts our attention to another special case -- one in which both informers observe the same signal. Both restrictions are strong enough to make the theorems inapplicable to most of the economic settings. Therefore we will deal with a much more realistic case below, in which both of the above restrictions are to large extent relaxed.

In most economic settings, almost all the informers available to the decision maker have more or less the same "type" of verifiability structures. In fact, there may only be a handful different "types" of verifiability structures in real life economic problems. Tirole in his seminal paper (Tirole, 1986) describes but one of the relatively common "types:" A "type" that the informer can either observe the true state of nature and prove that he does have observe it, or observe "nothing" and hence cannot prove or disprove anything. Similarly, for any quality checkers, it is much easier to verify that there is at least one defect in a certain design than to verify that the design is perfect, though different quality checker may have very different chances to detect any defect of the design, that is they observe very different signals of the quality of the design.

We have a conjecture that if the available informers are all of the same "type," the more Ω-informative is the signal observed by an informer, the more v-informative is the informer. If this conjecture is true, the applicability of the Blackwell theorem will be extended to a very large extent.

However, we encounter much difficulty in proving this conjecture. The first obstacle is in formalizing the concept of the "type" of a verifiability structure. When we say that two quality checkers, say John and May, are of the same "type," we have implicitly linked up the value "there are defects" in John's signal to the value "there are defects" in May's signal, and the value "perfect" in John's signal to the value "perfect" in May's signal. While this link seems natural in the example of quality checkers, it may seem too arbitrary when we are examining any two signals in abstract forms. In any way, the names of different values of a signal are assigned arbitrarily and therefore should not play any role in any economically meaningful comparison of informers. Therefore a prerequisite of formalizing the concept of the "type" of a verifiability structure is to understand the way people link up values of different signals.

It seems immediate that in at least one case this linking-up process can be understood more easily -- the case in which the two signals concerned are numerically equivalent. It is because when two signals are numerically equivalent, we can then view the "link" concerned simply as an one-to-one mapping between the signals. The next question is which one-to-one mapping can capture the essence of the "link" concerned.

Again, this one-to-one mapping can be most easily found in a further restricted case -- the case in which the signals are proxies of the state of nature. It means that for any signal Ω, and for any state of nature $\theta_i \in \Theta$, there is a corresponding $\omega_j \in \Omega$ such that people will be more confident to believe that the true state of nature is θ_i when ω_j has realized than when any other values of Ω have realized.[12] Formally we have:

---

12 In the example of quality checker, it will mean that people are more willing to believe that the true state of a commodity is "perfect" when the quality checker observes "perfect" than when he observes "there are defects."

Definition Ω is a *proxy* of Θ if:

(1) Ω is numerically equivalent to Θ; and

(2) there exists a one-to-one mapping, *f*, from Ω to Θ, such that $\forall \omega_j \in \Omega$, $\pi\left(f(\omega_j)|\omega_j\right) > \pi\left(f(\omega_j)|\omega_{k \neq j}\right)$.

Example Θ is a proxy of Θ.

If two signals are both proxies of Θ, there then exists an apparent way to link up the values of them. We first re-order the values of each signal (remember that the names given to each value are arbitrarily assigned, so any re-ordering should be allowed) according to the order of the states of nature. We called the re-ordered signal a well-ordered proxy of Θ.

Definition Ω is a *well-ordered* proxy of Θ if:

(1) Ω is numerically equivalent to Θ; and

(2) $\forall i \in I$, $\pi(\theta_i|\omega_i) > \pi(\theta_i|\omega_{j \neq i})$.

It means the i-th value should best predict the occurrence of the i-th state of nature. After we re-ordered the values of each signal, the natural way to link up the values of different signals is to link the i-th value of a signal to the i-th value to another signal.

Having the "link" formalized, the formalization of the concept of the "type" of verifiability structure is immediate.

Definition $(\Omega_1, \mathbf{V}_1)$ and $(\Omega_2, \mathbf{V}_2)$ are *of the same type* if:

(1) Ω_1 and Ω_2 are well-ordered proxies of Θ; and

(2) there exist a one-to-one mapping, *g*, from $\mathbf{V}_1$ to $\mathbf{V}_2$, such that $\forall V \in \mathbf{V}_1$, $g(V) = \bigcup_{\omega_j \in V} h(\omega_j)$, where h is a one-to-one mapping from Ω_1 to Ω_2 such that $h(\text{the j-th element of } \Omega_1) = \text{the j-th element of } \Omega_2$.

We can then formally present our conjecture as below.

Conjecture If $(\Omega_1, \mathbf{V}_1)$ and $(\Omega_2, \mathbf{V}_2)$ are of the same type, then $(\Omega_1, \mathbf{V}_1)$ is at least as v-informative as $(\Omega_2, \mathbf{V}_2)$ if Ω_1 is more Ω-informative than Ω_2.

Certainly we do not think we have formalized this conjecture satisfactorily. Remember that our formalization of the "link" between values of different signals as well as the "type" of a verifiability structure has excluded many interesting cases. We hope that further researches can fill the gap in this subject matter.

Moreover, even in this restricted case, we can only prove this conjecture in the special case that $|\Theta|=2$, where $|\Theta|$ is the cardinal number of Θ, and both the informer and the agent are risk-neutral. Again we hope the future researches can fill the blank space left untreated by this thesis.

The prove is presented in the following theorem.

Theorem 1.2 If $|\Theta|=2$, (Ω_1, V_1) and (Ω_2, V_2) are of the same type, and the informers as well as the agent are risk-neutral, then (Ω_1, V_1) is at least as v-informative as (Ω_2, V_2) if Ω_1 is more Ω-informative than Ω_2.

Proof By the Blackwell theorem, Ω_1 is more Ω-informative than Ω_2 if and only if there exists a Markov matrix, $\mathbf{M}$, such that $\mathbf{L}(\Omega_2) = \mathbf{L}(\Omega_1)\mathbf{M}$. Denote the typical elements of Ω_1, Ω_2 and $\mathbf{M}$ by ω_j, $\hat{\omega}_j$ and m_{ij} respectively.

Since both Ω_1 and Ω_2 are well-ordered proxies of Θ, we have:

$$\pi(\theta_1|\hat{\omega}_1) > \pi(\theta_1|\hat{\omega}_2)$$

$$\Rightarrow \frac{\pi(\theta_1)\pi(\hat{\omega}_1|\theta_1)}{\pi(\theta_1)\pi(\hat{\omega}_1|\theta_1)+\pi(\theta_2)\pi(\hat{\omega}_1|\theta_2)} > \frac{\pi(\theta_1)\pi(\hat{\omega}_2|\theta_1)}{\pi(\theta_1)\pi(\hat{\omega}_2|\theta_1)+\pi(\theta_2)\pi(\hat{\omega}_2|\theta_2)}$$

$$\Rightarrow \frac{\pi(\theta_1)\pi(\hat{\omega}_1|\theta_1)}{\pi(\theta_2)\pi(\hat{\omega}_1|\theta_2)} > \frac{\pi(\theta_1)(1-\pi(\hat{\omega}_1|\theta_1))}{\pi(\theta_2)(1-\pi(\hat{\omega}_1|\theta_2))}$$

$$\Rightarrow \pi(\hat{\omega}_1|\theta_1) > \pi(\hat{\omega}_1|\theta_2)$$

$$\Rightarrow \pi(\omega_1|\theta_1)m_{11} + \pi(\omega_2|\theta_1)m_{21} > \pi(\omega_1|\theta_2)m_{11} + \pi(\omega_2|\theta_2)m_{21}$$

$$\Rightarrow m_{11} > m_{21}.$$

Similarly, $\pi(\theta_2|\hat{\omega}_1) < \pi(\theta_2|\hat{\omega}_2) \Rightarrow m_{12} < m_{22}$.

Let $(\hat{Q},\hat{\gamma})$ be the solution to the problem (A, P, Ω_2, V_2, R). Let (Q,γ) be a mixed strategy such that $Q(\omega_j) = \sum_k m_{jk}\hat{Q}(\hat{\omega}_k)$ and $\gamma(\omega_j) = (\hat{\gamma}(\hat{\omega}_1), m_{j1}; \hat{\gamma}(\hat{\omega}_2), m_{j2})$. It suffices for us to show that the decision maker in the problem (A, P, Ω_1, V_1, R) can achieve an expected payoff of at least $\wp(\Omega_2, V_2; A, P, R)$ by applying (Q,γ).

We first show that (Q,γ) satisfies all the (IIC's) constraints in (A, P, Ω_1, V_1, R). For all $\omega_j \in \Omega_1$, for all $\omega_\mu \in M_1(\omega_j)$,

$$Q(\omega_j) - Q(\omega_\mu) = \sum_k (m_{jk} - m_{\mu k})\hat{Q}(\hat{\omega}_k)$$

$$= (m_{jj} - m_{\mu j})\hat{Q}(\hat{\omega}_j) + (m_{j\mu} - m_{\mu\mu})\hat{Q}(\hat{\omega}_\mu)$$
$$= (m_{jj} - m_{\mu j})\left(\hat{Q}(\hat{\omega}_j) - \hat{Q}(\hat{\omega}_\mu)\right).$$

Since (Ω_1, V_1) and (Ω_2, V_2) are of the same type, $\hat{\omega}_\mu \in M_2(\hat{\omega}_j)$. Therefore we have:

$$Q(\omega_j) - Q(\omega_\mu) \geq (m_{jj} - m_{\mu j}) \max\left\{\max_{\theta_i}\left\{R(\hat{\gamma}(\hat{\omega}_\mu), \theta_i) - R(\hat{\gamma}(\hat{\omega}_j), \theta_i)\right\}, 0\right\}.$$

Since $m_{jj} - m_{\mu j} > 0$, we have:

$$Q(\omega_j) - Q(\omega_\mu) \geq \max\left\{\max_{\theta_i}\left\{(m_{jj} - m_{\mu j})\left[R(\hat{\gamma}(\hat{\omega}_\mu), \theta_i) - R(\hat{\gamma}(\hat{\omega}_j), \theta_i)\right]\right\}, 0\right\}$$
$$= \max\left\{\max_{\theta_i}\left\{\sum_k m_{\mu k} R(\hat{\gamma}(\hat{\omega}_k), \theta_i) - \sum_k m_{jk} R(\hat{\gamma}(\hat{\omega}_k), \theta_i)\right\}, 0\right\}$$
$$= \max\left\{\max_{\theta_i}\left\{R(\gamma(\omega_\mu), \theta_i) - R(\gamma(\omega_j), \theta_i)\right\}, 0\right\}$$
$$= \Delta_{j\mu}.$$

Therefore all the (IIC's) constraints in (A, P, Ω_1, V_1, R) are satisfied.

We then show that by applying (Q, γ), the decision maker can achieve an expected payoff exactly equals to $\wp(\Omega_2, V_2; A, P, R)$. The decision maker's expected payoff of applying (Q, γ) is:

$$\sum_{\omega_j} \pi(\omega_j) \sum_{\theta_i} \pi(\theta_i | \omega_j)\left[P(\gamma(\omega_j), \theta_i) - Q(\omega_j)\right]$$
$$= \sum_{\omega_j} \pi(\omega_j) \sum_{\theta_i} \pi(\theta_i | \omega_j) \sum_k m_{jk}\left[P(\hat{\gamma}(\hat{\omega}_k), \theta_i) - \hat{Q}(\hat{\omega}_k)\right]$$
$$= \sum_{\theta_i} \pi(\theta_i) \sum_k \sum_j \pi(\omega_j | \theta_i) m_{jk}\left[P(\hat{\gamma}(\hat{\omega}_k), \theta_i) - \hat{Q}(\hat{\omega}_k)\right]$$
$$= \sum_{\theta_i} \pi(\theta_i) \sum_k \pi(\hat{\omega}_k | \theta_i)\left[P(\hat{\gamma}(\hat{\omega}_k), \theta_i) - \hat{Q}(\hat{\omega}_k)\right]$$
$$= \wp(\Omega_2, V_2; A, P, R).$$

Since it is true for all A, P and R, we conclude that (Ω_1, V_1) is at least as v-informative as (Ω_2, V_2). **Q.E.D.**

Section five: Conclusion

In this chapter we have taken the first step in comparing the "informativeness" of signals provided by some informers. Our work is important because we have explicitly deal with the possibility that the informers may cheat, which has long been ignored in the existing literature of decision making under uncertainty. We have introduced the idea of verifiability structure and the formal structure of the family of principal-informer-agent problems. These can be used as a corner-stone for future researches on the same subject matter. Moreover, we have drawn some preliminary insights in the comparison of the v-informativeness of informers. In particular, we prove that when two informers can observe the same signal, then one who can verify more is more v-informative than the other. We also prove that, in a restricted case, when informers are of the same "type," the Blackwell theorem is applicable in comparing the v-informativeness of informers.

Much work remains to be done. The concept of the "type" of a verifiability structure is still far from satisfactorily formalized. Moreover, the conjecture that "the Blackwell theorem is applicable in comparing the v-informativeness of informers whenever the informers are of the same 'type'" still waits for more thorough examination. Our work in this part is at most preliminary.

A real life economic setting that can apply the result of theorem 1.2 will be presented in chapter two, which deals with the Chinese central government's problem of signing responsibility contracts with its state enterprises.

CHAPTER TWO: IDIOSYNCRATIC OR UNIFORM CONTRACTS FOR THE CHINESE STATE ENTERPRISES

Section one: Introduction

In its reforming era, China gradually abandons its bureaucratic planning system over the state enterprises, and replaces it by various forms of contracting out systems. The common feature of these various forms of contracting out systems is to allow for greater autonomy for the enterprise managers. Under the reform, enterprise managers are supposed to be able to make their own decisions in areas like human resource management, production, marketing, investment..., etc. Though the autonomy is still far from perfect, and management is still subject to political interventions frequently, enterprise managers in general have much fewer obligations to fulfill. All remaining obligations like the amount of profit and tax to be submitted to the government are included in the responsibility contracts between the government and the enterprises. The decentralization of authority down to the hands of enterprise managers is recognized by most economists to be socially desirable, and further reform in this direction is frequently urged. However, it is impossible for the central government to conclude all the responsibility contracts with its subordinatory state enterprises.[13] It has to assign some subordinatory governmental units to do the job for it. One of the most usual ways is to simply assign the supervisory units of the enterprises to do so. In this case, the choice of supervisory units coincides with the choice of supervisors.[14,15] Yet sometimes the central government may find it more desirable for some major terms to be determined uniformly by higher governmental units, say industrial ministries, and to leave only minimal discretion for the supervisory units. In this case the choice of supervisors is separable from the choice of supervisory units.

13 There are as many as 104,400 in 1990 (State Statistical Bureau 1991, p.15).

14 The term "supervisor" used here differs from "supervisory unit" in the literature of Chinese economy. We use "supervisor" to denote the governmental unit which plays a dominant role in determining the terms of an enterprise's responsibility contract, while "supervisory unit" denotes the governmental unit to which the enterprise in question affiliates.

15 The term "supervisor" is used after Tirole's (1986) principal-supervisor-agent problem, in which a supervisor is one who inform the principal about a signal of the type of the agent.

The choice of suitable supervisors to represent the government to determine the terms of responsibility contracts of the enterprises seems to be one of the major unresolved problems of the Chinese government. While the governmental units high in the hierarchy clearly have much less local information about the specific environments of the enterprises under their control, governmental units down in the hierarchy may more easily collude with the enterprises they are bargaining and concluding responsibility contracts with. The difficulty of the choice is manifested by the unstable policy on this matter in the whole reform era. Table 1 tabulates the major changes in this regard.

Table 1

1978	enterprise fund (*qiye jijin*) system
1980	profit retention (*lirun liucheng*) system
1981	profit contract (*lirun baogan*) system
1983	tax-for-profit (*li gai shui*)
1987	management contract (*chengbao jingying zeren*) system

Under the enterprise fund system, enterprises were able to retain a part of the profit if they managed to fulfilled some obligatory targets. The size of the retained fund was a function of the set of targets the enterprise managed to fulfill. Since the function was uniform over different enterprises, little discretion power was left to the enterprises' supervisory units. The system thus approximated the extreme case where no local information had ever been called upon when the terms of contracts were determined.

The enterprise fund system was soon displaced by the profit retention system. The new system, while rationalizing of the old profit retention function, retained one of the major shortcomings of the old system: Local information was still not utilized. The result was foreseeable: Large variation in retained profits across different enterprises arose. In particular, enterprises which could more readily access planned material supplies or were producing products with high planned prices were enjoying unjustifiably high retained profits. Local information was proved to be especially important at the beginning of the reform when the profitability was still largely out of the control of the enterprise managers. The original profit retention system was soon modified into one where different contracts were designed for different industries. More local information was therefore utilized in the modified system.

The profit contract system represented a further development in the utilization of local information.[16] Idiosyncratic profit contract was concluded with every enterprise by its supervisory unit. But the employment of the more informed supervisors unnecessarily implied that the state could therefore extract more revenue from those more plannedly profitable enterprises. Du and Zhang (1992), among others, observe that "profit targets were generally set at inappropriately low levels [in the profit contracts]" (p.153). Chai (1991) also reports that under the system, "enterprises were successively provided with an increased share in profit" at the expense of state revenues, which in turn forced the state "to cut back on its own investment in such strategic sectors as transportation and energy" (p.51).

Partially due to the desire to cope with this problem, the Chinese government put forward the tax-for-profit system in 1983. Under the new system, a to large extent uniform tax system was applied to all enterprises. In fact, one of the rationales behind the substitution of tax for profit as the major form in which the enterprises submitted their profit to the state was that as the tax rates were fixed by law, it was hoped that less discretion power would be left for the enterprises' supervisory units. The large variation of profitability which eventually brought down the early version of the profit retention system was supposed to be coped with by the imposition of the adjustment tax. According to the tax law, the adjustment tax rate was a well defined and uniformly applicable function of the enterprise's 1983 profit level (Dong, 1988).

Clearly the original plan of the state was to level the retained profits without calling for the help of local informers. However, in practice the adjustment tax rates were never so determined, but were jointly set by the enterprises' supervisory units and the corresponding tax departments instead. The deformation of the original design of the adjustment tax could be

16 The profit contract system differed from the profit retention system also in the way the state shared profits with the enterprises. Under the profit contract system, enterprises were supposed to bear all the risk of running the business, though practically the persistence of soft budget constraint rescued most enterprises from bearing the risk of making loss.

attributed partially to the weakness of the state capacity, and partially to the sub-optimality of the original design.

The management contract system which has been implemented since 1987 represents the nominal return (while the real return had already happened after the deformation of the original design of the adjustment tax) of the idiosyncratic incentive system similar to the profit contract system. This nominal retreat from the ideal tax-for-profit system has been criticized by many economists who believe that a standard tax system is the correct direction of reform in the long run.[17] Wong (1987), among others, argues that the existing system "is particularly subject to abuse at the lower levels, where local officials see this as an opportunity to 'rob' the state treasury by reducing total tax payments" (p.397).

It seems that if governmental units down in the hierarchy have more local information about the enterprises, choosing a governmental unit that has more local information about the enterprises is not always desirable, or to speak in technical terms, *choosing an informer who can observe a more Ω-informative signal is not always desirable in practice.*

However, with the exception of chapter one in this thesis, there has been no attempt so far in the economic literature to explain the underlying reason of the above phenonemon. Worse still, the extremely idiosyncratic nature of the management contract system is still applauded at least by radical reformists, notwithstanding the abundant counter-evidences. In any way, it is reasonably natural for one to conjecture that a decision maker can never be worse off with an informer who can observe a more Ω-informative signal, for the decision maker can always duplicate what he will do when he faces an informer who can observe a less Ω-informative signal (see theorem 1.2 in chapter one and proposition 2.9 in section eight of this chapter).

To the best of my knowledge, this chapter is the first attempt to explain formally the mechanism underlying the phenomenon described above. Section two will set up the model of picking the appropriate supervisor for the state enterprises. In section three to seven, we will analysis the optimal contracts between the central government and the mid-tier governmental

17 See, for example, Wu and Liu (1991).

unit and the enterprise under different situations. The last section will compare the values of different mid-tier governmental units to the central government.

Section two: The model

There are three parties in the model: The central government, the mid-tier governmental unit and the state enterprise. We assume that all of them are risk neutral, so that we do not need to share our attention with the problem of risk sharing.

The enterprise produces a good whose value to the central government is normalized to 1. Denote the enterprise's output by y, and its cost of producing y units of the good by $\Psi(y,\Theta)$, where Θ is a productivity parameter of the enterprise. The cost increases with y and decreases with Θ: $\Psi_y > 0$, $\Psi_\Theta < 0$.[18] To guarantee that the second order conditions in the subsequent discussion automatically fulfilled, we make four more technical assumptions: $\Psi_{yy} > 0$, $\Psi_{yyy} > 0$, $\Psi_{y\Theta} < 0$ and $\Psi_{yy\Theta} < 0$. The first two assumptions say that the cost function is not only convex in y, but the convexity is increasing in y as well. The third and the fourth assumptions say that productive enterprise has a lower marginal cost as well as a smaller convexity in the cost function at every level of output.

For simplicity, we assume there are only two types of enterprises: The productive ones and the unproductive ones. It means that the productivity parameter Θ has only two possible values, namely $\bar{\theta}$ (productive) and $\underline{\theta}$ (unproductive). Correspondingly, we present $\Psi(y,\bar{\theta})$ as $\overline{\Psi}(y)$, and $\Psi(y,\underline{\theta})$ as $\underline{\Psi}(y)$. The prior probability of any enterprise to be productive is ν, and therefore that of unproductive is $1-\nu$.

Though the enterprises know their productivity well, the central government has absolutely no idea about it. However, the central government knows that the mid-tier governmental unit can observe a signal Ω of the enterprises' productivity Θ, and it can make use of this valuable local information by assigning the authority of concluding contracts down to the hand of the mid-tier governmental unit. Since analytically it is equivalent to concluding

---

18 Ψ_y and Ψ_Θ are the first derivatives of the cost function with respect to y and Θ respectively. The second and third derivatives will be similarly denoted in the subsequent discussion.

contracts by the central government itself under the advice of the mid-tier governmental unit, so it is suffice for us to model the process with a principal-supervisor-agent model. Moreover, according to our discussion in chapter one, the model can also be viewed as an informer game.

We assume that the signal observed by the mid-tier governmental unit has only two values: $\Omega = \{\omega_1, \omega_2\}$. It means that there can only be four types of verifiability structures:

type I: $\{\phi, \Omega\}$,

type II: $\{\phi, \{\omega_1\}, \Omega\}$,

type III: $\{\phi, \{\omega_2\}, \Omega\}$, and

type IV: $\{\phi, \{\omega_1\}, \{\omega_2\}, \Omega\}$.

A type I informer can verify none of the values he observes. A type II informer can verify an ω_1-report if he has observed ω_1. Similarly, a type III informer can verify an ω_2-report if he has observed ω_2. A type IV informer can verify any claim he makes to the central government.

The relationship between Θ and Ω can be fully described by two parameter. The first one is the prior probabilities of observing ω_1, denoted by n. The second is the conditional probability of an enterprise to be productive if ω_1 is observed, denoted by m. Therefore for any given signal, the posterior probabilities can be expressed in terms of v, n and m only:

$$\text{Prob}\ (\Theta = \bar{\theta} \mid \Omega = \omega_1) = m,$$
$$\text{Prob}\ (\Theta = \bar{\theta} \mid \Omega = \omega_2) = \frac{v - nm}{1-n},$$
$$\text{Prob}\ (\Theta = \underline{\theta} \mid \Omega = \omega_1) = 1\text{-}m,$$
$$\text{Prob}\ (\Theta = \underline{\theta} \mid \Omega = \omega_2) = \frac{1 - v - n(1-m)}{1-n}.$$

As the assignment of the names ω_1 and ω_2 to the two values of the signal is merely arbitrary, so without loss of generality we can always assume that $m > v$. That is, we always assume that ω_1 is a favorable signal of $\bar{\theta}$, and ω_2 is a favorable signal of $\underline{\theta}$.

Our next step is to capture the idea that the lower is the rank of the mid-tier governmental unit in question, the more Ω-informative is the signal observed by it. Suppose there is two mid-tier governmental units, with one higher in the hierarchy than another. Denote the signal observed by the high rank unit by Ω (n , m), and that of the low rank unit by

Ω (n' , m'). Then the idea that Ω (n' , m') is more Ω-informative than Ω (n , m) requires that there exist a Markov matrix **M** such that

(1) $$\mathbf{L}(\Omega(n, m)) = \mathbf{L}(\Omega(n', m'))\,\mathbf{M},$$

where $\mathbf{L}(\Omega(n, m))$ and $\mathbf{L}(\Omega(n', m'))$ are the likelihood matrices of Ω (n , m) and Ω (n' , m') respectively:

$$\mathbf{L}(\Omega(n,m)) = \begin{bmatrix} \dfrac{\nu - nm}{\nu} & \dfrac{nm}{\nu} \\ \dfrac{1-\nu-n(1-m)}{1-\nu} & \dfrac{n(1-m)}{1-\nu} \end{bmatrix},$$

$$\mathbf{L}(\Omega(n',m')) = \begin{bmatrix} \dfrac{\nu - n'm'}{\nu} & \dfrac{n'm'}{\nu} \\ \dfrac{1-\nu-n'(1-m')}{1-\nu} & \dfrac{n'(1-m')}{1-\nu} \end{bmatrix}.$$

Let the Markov matrix that satisfies (1) be

$$\mathbf{M} = \begin{bmatrix} c & 1-c \\ d & 1-d \end{bmatrix}.$$

Therefore the right hand side of (1) becomes

$$\begin{bmatrix} \dfrac{c\nu-(c-d)n'm'}{\nu} & \dfrac{(1-c)(\nu-n'm')+(1-d)n'm'}{\nu} \\ \dfrac{c(1-\nu)-(c-d)n'(1-m')}{1-\nu} & \dfrac{(1-c)(1-\nu-n'(1-m'))+(1-d)n'(1-m')}{1-\nu} \end{bmatrix}.$$

For (1) to hold, we must therefore have

(2) $$c\nu-(c-d)n'm' = \nu - nm, \text{ and}$$

(3) $$c\big(1-\nu-n'(1-m')\big)+dn'(1-m') = 1-\nu-n(1-m).$$

Solving the simultaneous equations (2) and (3), we have

(4) $$c = 1-\frac{n(m'-m)}{m'-\nu}, \text{ and}$$

(5) $$n'm'd = n'm'-nm+\left(\frac{n(m'-m)}{m'-\nu}\right)(\nu-n'm').$$

The definition of Markov matrix requires that c, d $\in$ [0 , 1]. For c to be less than one, we must have the second term on the right hand side of (4) to be positive and smaller than unity. As we always assume that $m' > \nu$, we must have $m' > m$. Given $m' > m$, the assumption that $m > \nu$ in turn ensures that $c > 0$. On the other hand, by transforming (5), we can see that the $d < 1$ is also automatically guaranteed whenever $m' > m$. Further calculation will yield the result that $d > 0$ if and only if $n' > \dfrac{n(m-\nu)}{c(m'-\nu)}$.

As a summary, Ω (n' , m') is more Ω-informative than Ω (n , m) if and only if m' > m and $n' > \frac{n(m-\nu)}{c(m'-\nu)}$. Since

$$\frac{m-\nu}{c(m'-\nu)} = \frac{m-\nu}{m'-\nu-n(m'-m)}$$
$$= \frac{m-\nu}{(m-\nu)+(m'-m)(1-n)} < 1,$$

it suffices for us to assume that $\begin{pmatrix} m' \\ n' \end{pmatrix} \geq \begin{pmatrix} m \\ n \end{pmatrix}$ in order to capture the idea that lower rank governmental unit can observe a more Ω-informative signal than higher rank governmental unit.

The timing of this principal-supervisor-agent game is as follows:

Period 1: The enterprise observes its type (productive or unproductive). The mid-tier governmental units observe their signals (ω_1 or ω_2).

Period 2: The central government chooses one of the available mid-tier governmental units as its informer.

Period 3: The central government signs (responsibility) contracts with its informer and the enterprise respectively. The enterprise signs bribery contract with the mid-tier governmental unit.

Period 4: The mid-tier governmental unit report what it has observed to the central government. The enterprise produces certain amount of output. The central government rewards the mid-tier governmental unit and the enterprise according to the (responsibiliy) contracts. The enterprise pays the bribe to the mid-tier governmental unit according to the bribery contract. Denote the reward given by the central government to the mid-tier governmental unit by Q, that to the enterprise by R, and the bribe given by the enterprise to the mid-tier governmental unit by B.

We assume that whenever the enterprise gives one dollar to the mid-tier governmental unit, the latter will only receive $k \in [0, 1]$ dollar. The reciprocal of k can be viewed as an index of the likelihood that the bribery is to be detected. Since there is always risk incurred in the bribery process, one dollar paid out of the enterprise's pocket will be perceived as less than one dollar in the eye of the mid-tier governmental unit.[19]

As the last step of specifying the model, we have to specify the objective functions of the parties involved. We first normalize the enterprise's and the mid-tier governmental unit's reservation returns to zero. We then assume that the enterprise's objective is to maximize R-Ψ-B; the mid-tier governmental unit's objective is to maximize Q+B. The central government's objective function is a weighted summation of its own return, the mid-tier governmental unit's return, and the enterprise's return: $W = y - R - Q + \lambda^E (R - \Psi - B) + \lambda^I (Q + B)$. Here λ^E and λ^I are the weight of the enterprise's return and the mid-tier governmental unit's return in the central government's objective function respectively. If they both equal to one, then the central government is benevolent. If they both equal to zero, then the central government only cares about its own revenue. In general we assume that they both lie at somewhere between zero and one.

Section three: Complete information case

As a benchmark, we first investigate how will the central government conclude responsibility contract with an enterprise if the central government has complete information of the enterprise's productivity. For expositional purpose, we confine our attention to the class of linear incentive schemes only (as we will do in the rest of this chapter), that is, we only consider incentive schemes of the form R = a + by. Under complete information, there is no need for any help from the mid-tier governmental unit, and the optimal contract in this case has already been worked out by Freixas, Guesnerie and Tirole (1985). The central government's problem is:

(6) $$\max_{a,b,y} W = y - \lambda^E \Psi(y) - (1 - \lambda^E)(a + by),$$

subject to the enterprise incentive compatibility constraint:

(7) $$y \in \arg\max_{y'} a + by' - \Psi(y'),$$

and the enterprise individual rationality constraint:

---

19 For an interpretation of k>1, see Tirole (1992).

(8) $a + by - \Psi(y) \geq 0$.

Rewrite (7) as $b = \Psi'(y)$,[20] and incorporate it as well as (8) into (6). The central government's problem becomes:

(6') $\max_b W = y(b) - \Psi(y(b)),$

where y(.) denotes the choice of y made by the enterprise when the slope of the contract is b. The first order condition of (6') is b = 1. The second order condition is automatically satisfied given the technical assumptions specified at the beginning of section two.[21] Denote y(1) by y*. Therefore we have:

Proposition 2.1 Under complete information, the optimal incentive scheme for whatever type of enterprise is $R = \Psi(y^*) - y^* + y$.

Note that unity is also the value of the product produced by the enterprise, so an incentive scheme with slope equals to unity also maximizes the social welfare.

Section four: Incomplete information with type I informer

From this section on we will investigate into the cases of incomplete information. We will first consider the case of incomplete information with type I informer in this section. Recall that a type I informer has the following verifiability structure: $\{\phi, \Omega\}$. That is, a type I informer cannot verify any claim he makes. As we have argued in chapter one, a decision maker can never do better with a type I informer than with no informer at all. So this case is also equivalent to that of no informer.

Appealing to the revelation principle (Myerson, 1979), the central government cannot do better than applying the optimal direct mechanism. So it suffices to consider the class of direct mechanisms only. Suppose the central government offers the enterprise two kinds of

---

20 It is the first order approach to the principal-agent problem (see Rogerson, 1985), which is unnecessarily legitimate. We are legitimate to do so because the enterprise's payoff function is strictly concave in the output level, so any stationary point is by itself global maximum.

21 The second order condition of (6') is $y''(b) - \Psi_{yy}(y)(y'(b))^2 - \Psi_y(y)y''(b) \leq 0$. By the technical assumptions, Ψ_{yy} is positive, so the second term is negative. From (7) and the first order condition, we know that $\Psi_y = b = 1$, so the first and the third terms can be cancelled out. Therefore the second order condition is satisfied.

incentive schemes, contingent on the type the enterprise reports it belongs to. The incentive scheme offered to the reportedly productive enterprise is $\overline{R} = \bar{a} + \bar{b}\bar{y}$, and that for the reportedly unproductive one is $\overline{R} = \underline{a} + \underline{b}\underline{y}$. The central government's problem becomes:

(9) $$\max_{\bar{a},\bar{b},\bar{y},\underline{a},\underline{b},\underline{y}} \; W = v\left[\bar{y} - \lambda^E \overline{\Psi}(\bar{y}) - (1-\lambda^E)(\bar{a} + \bar{b}\bar{y})\right] + (1-v)\left[\underline{y} - \lambda^E \underline{\Psi}(\underline{y}) - (1-\lambda^E)(\underline{a} + \underline{b}\underline{y})\right],$$

subject to four incentive compatibility constraints:

(10) $$\overline{\Psi}'(\bar{y}) = \bar{b},$$

(11) $$\underline{\Psi}'(\underline{y}) = \underline{b},$$

(12) $$\bar{a} + \bar{b}\bar{y} - \overline{\Psi}(\bar{y}) \geq \underline{a} + \underline{b}\bar{y}(\underline{b}) - \overline{\Psi}(\bar{y}(\underline{b})),$$

(13) $$\underline{a} + \underline{b}\underline{y} - \underline{\Psi}(\underline{y}) \geq \bar{a} + \bar{b}\underline{y}(\bar{b}) - \underline{\Psi}(\underline{y}(\bar{b})),$$

and two individual rationality constraints:

(14) $$\bar{a} + \bar{b}\bar{y} - \overline{\Psi}(\bar{y}) \geq 0,$$

(15) $$\underline{a} + \underline{b}\underline{y} - \underline{\Psi}(\underline{y}) \geq 0,$$

where $\bar{y}(.)$ and $\underline{y}(.)$ are the choices of output levels made by the productive and unproductive enterprises respectively. (10) and (11) are simply the analogies of (7) for the productive and unproductive firms respectively. (12) and (13) guarantee that productive enterprises will not pretend to be unproductive, and unproductive ones will not pretend to be productive. (14) and (15) corresponds to (8) for the productive and unproductive enterprises.

As is standard in the literature, given (15), the individual rationality constraint (14) is never binding.[22] Since it will never be optimal for both (14) and (15) to be unbinding,[23] we conclude that (15) must be binding. As (14) is never binding, we see no reason why $\bar{a}$ should not be suppressed to as low a level as possible until (12) becomes binding. At last, (12) reduces constraint (13) to:

---

22 The logic is simple: The productive enterprise can always duplicate what the unproductive one would do. While it receives the same amount of reward as the unproductive enterprise does in this case, it will incur a smaller production cost. So it can always guarantee itself a reward larger than zero by doing exactly what the unproductive enterprise does.

23 If both (14) and (15) are unbinding, there must exist a positive amount such that the central government can lower $\bar{a}$ and $\underline{a}$ equally by that amount without affecting inequalities in (14) and (15). Note that by doing so the central government can improve its welfare, while the originally satisfied constraints (12) and (13) will remain satisfied.

(13') $\underline{b} \le \bar{b}$,[24]

which is actually unbinding. As we will see very soon, even if we deliberately ignore (13') when we solve the problem, the solution will automatically satisfy (13').

This reduces our total number of constraints down to only four. Put aside the two unbinding constraints (13) and (14), and incorporate the others into (9). The central government's problem can be transformed to:

(9') $$\max_{\bar{b},\underline{b}} W = \nu\Big[\bar{y}(\bar{b}) - \lambda^E \bar{\Psi}(\bar{y}(\bar{b})) - (1-\lambda^E)\big(\underline{\Psi}(\underline{y}(\underline{b})) - \underline{b}\underline{y}(\underline{b}) + \underline{b}\bar{y}(\underline{b}) - \bar{\Psi}(\bar{y}(\underline{b})) + \bar{\Psi}(\bar{y}(\bar{b}))\big)\Big] + (1-\nu)\big[\underline{y}(\underline{b}) - \underline{\Psi}(\underline{y}(\underline{b}))\big].$$

Solving the problem yields the following two first order conditions:

(16) $$\bar{y}'(\bar{b}) = \bar{b}\bar{y}'(\bar{b}), \text{ and}$$

(17) $$1-\underline{b} = \frac{\nu(1-\lambda^E)\big(\bar{y}(\underline{b}) - \underline{y}(\underline{b})\big)}{(1-\nu)\underline{y}'(\underline{b})}.$$

Condition (16) implies that $\bar{b} = 1$. The right hand side of (17) is positive, implying that $\underline{b} < 1 = \bar{b}$. The second order conditions are guaranteed by the technical assumptions introduced in section two. The result is standard in the literature: The unproductive enterprise is to produce at below first best level so that the central government can save some money in motivating the productive one to produce at the first best level (Arrow, 1986).

Proposition 2.2 Under incomplete information with type I informer, the optimal contract is $\underline{R} = \underline{\Psi}(\underline{y}) - \underline{b}\underline{y} + \underline{b}y$ for unproductive enterprise and

---

24 **Proof** Adding (12) to (13), we have:

$$\Big[\big(\bar{\Psi}(\bar{y}(\bar{b})) - \bar{b}\bar{y}(\bar{b})\big) - \big(\bar{\Psi}(\bar{y}(\underline{b})) - \underline{b}\bar{y}(\underline{b})\big)\Big]$$
$$-\Big[\big(\underline{\Psi}(\underline{y}(\bar{b})) - \bar{b}\underline{y}(\bar{b})\big) - \big(\underline{\Psi}(\underline{y}(\underline{b})) - \underline{b}\underline{y}(\underline{b})\big)\Big] \le 0$$
$$\Rightarrow \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{b}}^{\bar{b}} \frac{\partial^2}{\partial\Theta\partial b}\big(\Psi(y(b)) - by(b)\big) db d\Theta \le 0$$
$$\Rightarrow \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{b}}^{\bar{b}} \frac{\partial}{\partial\Theta}\big(\Psi_y y'(b) - y(b) - by'(b)\big) db d\Theta \le 0$$
$$\Rightarrow \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{b}}^{\bar{b}} \left(-\frac{\partial}{\partial\Theta} y(b)\right) db d\Theta \le 0$$
$$\Rightarrow \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{b}}^{\bar{b}} \frac{\Psi_{y\Theta}}{\Psi_{yy}} db d\Theta \le 0$$
$$\Rightarrow \bar{b} \ge \underline{b}.$$ **Q.E.D.**

$\overline{R} = \underline{\Psi}(\underline{y}) - \underline{b}\underline{y} + \underline{b}\overline{y}(\underline{b}) - \overline{\Psi}(\overline{y}(\underline{b})) + \overline{\Psi}(y^*) - y^* + y$ for productive one, where y* is defined in section two, and $\underline{b}$, $\overline{y}(.)$ and $\underline{y}(.)$ in equations (10), (11) and (17). The informer will not be called into service and thus only receives his reservation wage.

Before we end this section, we would like to derive the following lemma which will be useful in the subsequent sections.

Lemma 2.1 $\underline{b}$ decreases with ν.

Proof From the first order condition (17), we have:

$$\frac{d\underline{b}}{d\nu} = -\frac{-(1-\underline{b})\underline{y}'(\underline{b}) - (1-\lambda^E)(\overline{y}(\underline{b}) - \underline{y}(\underline{b}))}{\dfrac{\partial^2 W}{\partial \underline{b}^2}}.$$

By our technical assumptions, W is strictly concave in $\underline{b}$,[25] so the denominator on the right side is negative. The derivative is therefore negative. **Q.E.D.**

Lemma 2.1 says that the higher is the probability that an enterprise is unproductive, the closer is the slope of the incentive scheme offered to the unproductive enterprise to its first best level, as the cost of deviating from the first best level is also higher now.

Hereafter we will denote $\underline{b}$ by $\underline{b}(.)$ whenever we want to stress that $\underline{b}$ is a function of the probability of an enterprise to be productive.

Lemma 2.2 Holding the productive enterprise incentive compatibility constraint and the unproductive enterprise individual rationality constraint binding, $\overline{R} - \overline{\Psi}(\overline{y})$ increases with $\underline{b}$.

Proof $\dfrac{\partial}{\partial \underline{b}}\left[\overline{R} - \overline{\Psi}(\overline{y})\right] = \dfrac{\partial}{\partial \underline{b}}\left[\underline{a} + \underline{b}\overline{y}(\underline{b}) - \overline{\Psi}(\overline{y}(\underline{b}))\right]$

$$= \frac{\partial}{\partial \underline{b}}\left[\underline{\Psi}(\underline{y}(\underline{b})) - \underline{b}\underline{y}(\underline{b}) + \underline{b}\overline{y}(\underline{b}) - \overline{\Psi}(\overline{y}(\underline{b}))\right] = \overline{y}(\underline{b}) - \underline{y}(\underline{b})$$

$$= \int_{\underline{\theta}}^{\bar{\theta}} \left(-\frac{\Psi_{y\Theta}}{\Psi_{yy}}\right) d\Theta > 0. \quad \textbf{Q.E.D.}$$

---

25 **Proof** $\dfrac{\partial^2 W}{\partial \underline{b}^2} = \nu(1-\lambda^E)(\underline{y}'(\underline{b}) - \overline{y}'(\underline{b})) + (1-\nu)((1-\underline{b})\underline{y}''(\underline{b}) - \underline{y}'(\underline{b}))$

$$= \nu(1-\lambda^E)\int_{\underline{\theta}}^{\bar{\theta}} \frac{\Psi_{yy\Theta}(y(\underline{b}))}{(\Psi_{yy}(y(\underline{b})))^2} - (1-\nu)\left[\frac{(1-\underline{b})\Psi_{yyy}(y(\underline{b}))}{(\Psi_{yy}(y(\underline{b})))^2} + \frac{1}{\Psi_{yy}(y(\underline{b}))}\right]$$

< 0 **Q.E.D.**

Lemma 2.2 shows that the productive enterprise will be better off if the slope of the incentive scheme offered to the unproductive enterprise is closer to its first best level. Since we have already shown that this will be the case if the probability that an enterprise is productive is smaller in lemma 2.1, we can therefore conclude that the smaller the probability that an enterprise is productive, the better is the productive enterprise.

Section five: Incomplete information with type II informer

~~The central government continues to be ignorant about the enterprise's productivity.~~ But this time it is lucky enough to have a type II informer. Recall that a type II informer has the following verifiability structure: $\{\phi,\{\omega_1\},\Omega\}$. This type of informers can verify an ω_1-report but not an ω_2-report. With the help of the informer, the central government can offer up to four different incentive schemes to the enterprise contingent to the reported type of the enterprise and the report made by the informer:

$\overline{R}_1 = \overline{a}_1 + \overline{b}_1 y$ for productive enterprise when informer reports ω_1,

$\underline{R}_1 = \underline{a}_1 + \underline{b}_1 y$ for unproductive enterprise when informer reports ω_1,

$\overline{R}_2 = \overline{a}_2 + \overline{b}_2 y$ for productive enterprise when informer reports ω_2, and

$\underline{R}_2 = \underline{a}_2 + \underline{b}_2 y$ for unproductive enterprise when informer reports ω_2.

The central government's problem is:

$$
\begin{aligned}
(18)\quad \max_{\substack{\overline{a}_1,\overline{a}_2,\underline{a}_1,\underline{a}_2,\\ \overline{b}_1,\overline{b}_2,\underline{b}_1,\underline{b}_2}} \; W = {} & nm\left[\overline{y}(\overline{b}_1) - \lambda^E \overline{\Psi}(\overline{y}(\overline{b}_1)) - (1-\lambda^E)\overline{R}_1 - (1-\lambda^I)Q(\omega_1)\right] \\
& + n(1-m)\left[\underline{y}(\underline{b}_1) - \lambda^E \underline{\Psi}(\underline{y}(\underline{b}_1)) - (1-\lambda^E)\underline{R}_1 - (1-\lambda^I)Q(\omega_1)\right] \\
& + (\nu - nm)\left[\overline{y}(\overline{b}_2) - \lambda^E \overline{\Psi}(\overline{y}(\overline{b}_2)) - (1-\lambda^E)\overline{R}_2 - (1-\lambda^I)Q(\omega_2)\right] \\
& + (1-\nu-n(1-m))\left[\underline{y}(\underline{b}_2) - \lambda^E \underline{\Psi}(\underline{y}(\underline{b}_2)) \right.\\
& \qquad \left. - (1-\lambda^E)\underline{R}_2 - (1-\lambda^I)Q(\omega_2)\right],
\end{aligned}
$$

where the maximand is the expected payoff of the central government.

In order to guarantee that, no matter what the informer reports, both types of enterprise will not pretend to be the other types, we have the following enterprise incentive compatibility constraints on the incentive schemes:

$$
(19)\qquad \overline{a}_1 + \overline{b}_1 \overline{y}(\overline{b}_1) - \overline{\Psi}(\overline{y}(\overline{b}_1)) \geq \underline{a}_1 + \underline{b}_1 \overline{y}(\underline{b}_1) - \overline{\Psi}(\overline{y}(\underline{b}_1)),
$$

(20) $$\bar{a}_2 + \bar{b}_2 \bar{y}(\bar{b}_2) - \bar{\Psi}(\bar{y}(\bar{b}_2)) \geq \underline{a}_2 + \underline{b}_2 \bar{y}(\underline{b}_2) - \bar{\Psi}(\bar{y}(\underline{b}_2)),$$

(21) $$\underline{a}_1 + \underline{b}_1 \underline{y}(\underline{b}_1) - \underline{\Psi}(\underline{y}(\underline{b}_1)) \geq \bar{a}_1 + \bar{b}_1 \underline{y}(\bar{b}_1) - \underline{\Psi}(\underline{y}(\bar{b}_1)), \text{ and}$$

(22) $$\underline{a}_2 + \underline{b}_2 \underline{y}(\underline{b}_2) - \underline{\Psi}(\underline{y}(\underline{b}_2)) \geq \bar{a}_2 + \bar{b}_2 \underline{y}(\bar{b}_2) - \underline{\Psi}(\underline{y}(\bar{b}_2)).$$

(19) and (20) are simply the ω_1- and ω_2-report versions of (12), while (21) and (22) are those of (13).

Similarly, we have four enterprise individual rationality constraints:

(23) $$\bar{a}_1 + \bar{b}_1 \bar{y}(\bar{b}_1) - \bar{\Psi}(\bar{y}(\bar{b}_1)) \geq 0,$$

(24) $$\bar{a}_2 + \bar{b}_2 \bar{y}(\bar{b}_2) - \bar{\Psi}(\bar{y}(\bar{b}_2)) \geq 0,$$

(25) $$\underline{a}_1 + \underline{b}_1 \underline{y}(\underline{b}_1) - \underline{\Psi}(\underline{y}(\underline{b}_1)) \geq 0, \text{ and}$$

(26) $$\underline{a}_2 + \underline{b}_2 \underline{y}(\underline{b}_2) - \underline{\Psi}(\underline{y}(\underline{b}_2)) \geq 0.$$

(23) and (24) are the ω_1- and ω_2-report versions of (14), while (25) and (26) are those of (15). As before, (25) and (26) make (23) and (24) unbinding. Moreover, (19) and (20) also reduce (21) and (22) to:

(21') $$\underline{b}_1 \leq \bar{b}_1, \text{ and}$$

(22') $$\underline{b}_2 \leq \bar{b}_2,$$

respectively. However, whether (25) and (26) are necessarily binding is not yet clear at this place.

The reason why we cannot tell immediately whether (25) and (26) are binding is that the situation is more complicated than our previous situation with type I informer. Now we know that our (type II) informer can verify ω_1 but not ω_2. So if the informer reports ω_2, the central government will hesitate to believe it. In particular if the enterprise's payoff is higher when the informer reports ω_2 rather than ω_1, an informer observing ω_1 will have every incentive to take bribery from the enterprise and then submit a false report of ω_2 to the central government. In order to motivate an informer observing ω_1 to report truly, the central government has to reward report of ω_1.[26] The minimum reward is the maximum bribery an

26 An informer observing ω_2 will always reports truely, as he cannot verify a report of ω_1.

informer observing ω_1 can receive if it reports ω_2. Thus we have the following informer incentive compatibility constraints:[27]

(27) $$Q(\omega_1) - Q(\omega_2) \geq k \cdot \max\left\{\left[\overline{R}_2 - \overline{\Psi}\left(\overline{y}(\overline{b}_2)\right)\right] - \left[\overline{R}_1 - \overline{\Psi}\left(\overline{y}(\overline{b}_1)\right)\right], \left[\underline{R}_2 - \underline{\Psi}\left(\underline{y}(\underline{b}_2)\right)\right] - \left[\underline{R}_1 - \underline{\Psi}\left(\underline{y}(\underline{b}_1)\right)\right], 0\right\}, \text{ and}$$

(28) $$Q(\omega_2) \geq 0.$$

The third term in the brace corresponds to the reward to report of ω_2, which equals to the reservation wage of the informer (zero by our assumption). It is clear that both constraints must be binding.

Since $\overline{R}_2$, $\underline{R}_2$ and $Q(\omega_1)$ are all (weakly) increasing in $\underline{a}_2$, there is no reason why $\underline{a}_2$ is not suppressed to as low a level as possible until (26) is binding. The fact that (26) is binding implies that the second term in the brace of (27) can never be positive. So (27) can be rewritten as:

(27') $$Q(\omega_1) = k \cdot \max\left\{\left[\overline{R}_2 - \overline{\Psi}\left(\overline{y}(\overline{b}_2)\right)\right] - \left[\overline{R}_1 - \overline{\Psi}\left(\overline{y}(\overline{b}_1)\right)\right], 0\right\}.$$

Similarly, both $\underline{R}_2$ and $Q(\omega_1)$ are (weakly) increasing in $\overline{a}_2$. As reducing $\overline{a}_2$ can always enhance the welfare of the central government constraint (20) must also be binding.

However, though $\overline{R}_2$ and $\underline{R}_2$ are (weakly) increasing in $\underline{a}_1$, $Q(\omega_1)$ is (weakly) decreasing in it. So suppressing $\underline{a}_1$ as low a level as possible may not always be desirable. The same is also true for $\overline{a}_1$. Therefore it is still not clear by now whether (19) and (25) are binding or not. Nevertheless, it is not that difficult to figure out the condition under which the central government will choose to reduce $\overline{R}_1$ and $\underline{R}_1$ rather than $Q(\omega_1)$.

Suppose the initial contract is characterized by an unbinding constraint (25). What will happen if the central government reduces both $\overline{a}_1$ and $\underline{a}_1$ by ε? If ε is small enough, all the originally unbinding constraints can remain unbinding.[28] Therefore the benefit of doing so is that with probability n (the probability of reporting ω_1), the central government will pay ε less

---

27 The introduction of constraint (27) and (28) implies that we are confining ourself to the set of collusion-proof contracts only. In this simple setting, the decision maker (the central government) cannot do better than offering the optimal collusion-proof contract. See Tirole (1986).

28 The changes on the left and right sides of constraints (19) and (21) will automatically cancel each other.

to the enterprise. The cost of doing so depends on whether the first term in the brace of (27') (hereafter denoted by Δ) is non-negative. If it is, the cost is that with the same probability, the central government has to pay $k\varepsilon$ more to the mid-tier governmental unit. If it is not, there will be no cost incurred at all. So the net gain from reducing $\bar{a}_1$ and $\underline{a}_1$ is at least $n[(1-\lambda^E)-(1-\lambda^I)k]\varepsilon$, which is independent of the original levels of $\bar{a}_1$ and $\underline{a}_1$. The central government will choose to reduce $\bar{a}_1$ and $\underline{a}_1$ if:

(29) $$(1-\lambda^E) \geq (1-\lambda^I)k.$$

If (29) does not hold, we immediately have the following results:

Lemma 2.3 $Q(\omega_1)=0$ if (29) does not hold.

Proof Suppose not, then Δ must be positive. The central government can gain from decreasing $Q(\omega_1)$ by $k\varepsilon$ and increasing both $\bar{a}_1$ and $\underline{a}_1$ by ε. **Q.E.D.**

Lemma 2.4 $\Delta = 0$ if (29) does not hold.

Proof Suppose not, by lemma 2.3, $\Delta < 0$. (25) must then be binding, or else the central government can always improve its own welfare by decreasing both $\bar{a}_1$ and $\underline{a}_1$ by ε. Similarly, (19) must be binding, or else the central government can always do better by decreasing $\bar{a}_1$ by ε. Binding constraints (19) and (25) reduce (21) to:

(21') $$\underline{b}_1 \leq \underline{b}_2.$$

Solving (18) subject to $Q(\omega_1)=Q(\omega_2)=0$ and binding constraints (19), (20), (25), (26),[29] we have the optimal contract characterized by $\bar{b}_1=\bar{b}_2=1$, $\underline{b}_1=\underline{b}(m)$, $\underline{b}_2=\underline{b}\left(\frac{\nu-nm}{1-n}\right)$, and $\bar{a}_1$, $\bar{a}_2$, $\underline{a}_1$ and $\underline{a}_2$ given by (19), (20), (25) and (26) respectively. By lemma 2.1, we have $\underline{b}_2>\underline{b}_1$, and by lemma 2.2, $\Delta>0$, which is contradictory to our initial assumption. **Q.E.D.**

Lemma 2.4 implies that:

(30) $$\bar{a}_1+\bar{b}_1\bar{y}(\bar{b}_1)-\bar{\Psi}(\bar{y}(\bar{b}_1))=\bar{a}_2+\bar{b}_2\bar{y}(\bar{b}_2)-\bar{\Psi}(\bar{y}(\bar{b}_2)).$$

Lemma 2.5 (19) must be binding if (29) does not hold.

---

29 (21') and (22') are not binding, as will be shown by the characterizations of the optimal contract.

Proof Suppose not, the central government can then gain by moving $\underline{b}_1$ towards unity, and adjusting $\underline{a}_1$ so that $\underline{R}_1 - \underline{\Psi}(\underline{y}(\underline{b}_1))$ can be held intact, until either (19) becomes binding (the proof is then finished) or $\underline{b}_1 = 1$. By equation (30) and binding constraints (20) and (26), the fact that (19) is unbinding necessarily implies that $\underline{b}_2 > \underline{b}_1$.[30] We therefore have $\underline{b}_2 > 1$, which cannot be optimal.[31] So if (29) does not hold, (19) must be binding. **Q.E.D.**

Lemma 2.6 (25) must be binding if (19) does not hold.

Proof Suppose not, then (21) must be binding, or else the central government can gain by lowering $\underline{a}_1$ by ε. The fact that both (19) and (21) are binding implies that $\bar{a}_1 = \underline{a}_1$ and $\bar{b}_1 = \underline{b}_1$. The central government is not going to offer separate incentive schemes for different types of enterprise when the informer reports ω_1. Solving the problem of (18) subject to $\bar{a}_1 = \underline{a}_1$, $\bar{b}_1 = \underline{b}_1$, (30), and the binding constraints (20) and (26),[32] we have three first order conditions:

$$\frac{\partial W}{\partial \bar{b}_1} = 0 \Rightarrow nm\left(\bar{y}'(\bar{b}_1) - \bar{b}_1 \bar{y}'(\bar{b}_1)\right)$$
$$+ n(1-m)\left[\underline{y}'(\bar{b}_1) - \bar{b}_1 \underline{y}'(\bar{b}_1) + (1-\lambda^E)\left(\bar{y}(\bar{b}_1) - \underline{y}(\bar{b}_1)\right)\right] = 0$$
$$\Rightarrow \bar{b}_1 - 1 = \frac{(1-m)(1-\lambda^E)\left(\bar{y}(\bar{b}_1) - \underline{y}(\bar{b}_1)\right)}{m\left(\bar{y}'(\bar{b}_1) - \underline{y}'(\bar{b}_1)\right) + \underline{y}'(\bar{b}_1)} > 0$$
$$\Rightarrow \bar{b}_1 > 1,$$
$$\frac{\partial W}{\partial \bar{b}_2} = 0 \Rightarrow (\nu - nm)\left(\bar{y}'(\bar{b}_2) - \bar{b}_2 \bar{y}'(\bar{b}_2)\right) = 0$$
$$\Rightarrow \bar{b}_2 = 1, \text{ and}$$
$$\frac{\partial W}{\partial \underline{b}_2} = 0 \Rightarrow (1-\nu-n(1-m))\left(\underline{y}'(\underline{b}_2) - \underline{b}_2 \underline{y}'(\underline{b}_2)\right)$$
$$-(\nu + n(1-m))(1-\lambda^E)\left(\bar{y}(\underline{b}_2) - \underline{y}(\underline{b}_2)\right) = 0$$
$$\Rightarrow 1 - \underline{b}_2 = \frac{(\nu + n(1-m))(1-\lambda^E)\left(\bar{y}(\underline{b}_2) - \underline{y}(\underline{b}_2)\right)}{(1-\nu-n(1-m))\underline{y}'(\underline{b}_2)} > 0$$
$$\Rightarrow \underline{b}_2 < 1.$$

30 $\underline{b}_2 = \underline{b}_1$ only in the limiting case where (19) and (25) are both binding.

31 The reason why $\underline{b}_2$ should not be larger than unity, that is should not be larger than the first best level, can be easily seen from the fact that increasing $\underline{b}_2$ and thus $\underline{y}(\underline{b}_2)$ will not only increase the production cost (which is the *only* cost incurred in the complete information case), but also increase the payment to the productive enterprise (via (20) and (30)).

32 Constraint (21') is not binding, as can be seen soon.

Combining the first and the third first order conditions, we have $\bar{b}_1 = \underline{b}_1 > \underline{b}_2$. However, the fact that (25) is not binding implies that $\underline{b}_1 \le \underline{b}_2$. Contradiction arises.

Q.E.D.

We now have ground to work out the optimal contract under the circumstance that (19) does not hold. By lemma 2.5 and 2.6, (19) and (25) are binding, so we have $\underline{b}_1 = \underline{b}_2$. Solving (18) subject to $\underline{b}_1 = \underline{b}_2$, (30) and binding constraints (20), (25) and (26),[33] we have the following three first order conditions:

$$\frac{\partial W}{\partial \bar{b}_1} = 0 \Rightarrow nm\left(\bar{y}'(\bar{b}_1) - \bar{b}_1 \bar{y}'(\bar{b}_1)\right) = 0$$

$$\Rightarrow \bar{b}_1 = 1,$$

$$\frac{\partial W}{\partial \bar{b}_2} = 0 \Rightarrow (\nu - nm)\left(\bar{y}'(\bar{b}_2) - \bar{b}_2 \bar{y}'(\bar{b}_2)\right) = 0$$

$$\Rightarrow \bar{b}_2 = 1, \text{ and}$$

$$\frac{\partial W}{\partial \underline{b}_1} = 0 \Rightarrow (1-\nu)\left(\underline{y}'(\underline{b}_1) - \underline{b}_1 \underline{y}'(\underline{b}_1)\right) - (1-\lambda^E)\nu\left(\bar{y}(\underline{b}_1) - \underline{y}(\underline{b}_1)\right) = 0$$

$$\Rightarrow 1 - \underline{b}_1 = \frac{(1-\lambda^E)\nu\left(\bar{y}(\underline{b}_1) - \underline{y}(\underline{b}_1)\right)}{(1-\nu)\underline{y}'(\underline{b}_1)} > 0$$

$$\Rightarrow \underline{b}_1 < 1.$$

Note that the optimal contract does not distinguish the incentive schemes under ω_1- and ω_2-reports. This means that the informer has not been called into service under the optimal contract. Therefore we have:

---

33 (21') and (22') are not binding at the optimal solution, as will become clear very soon.

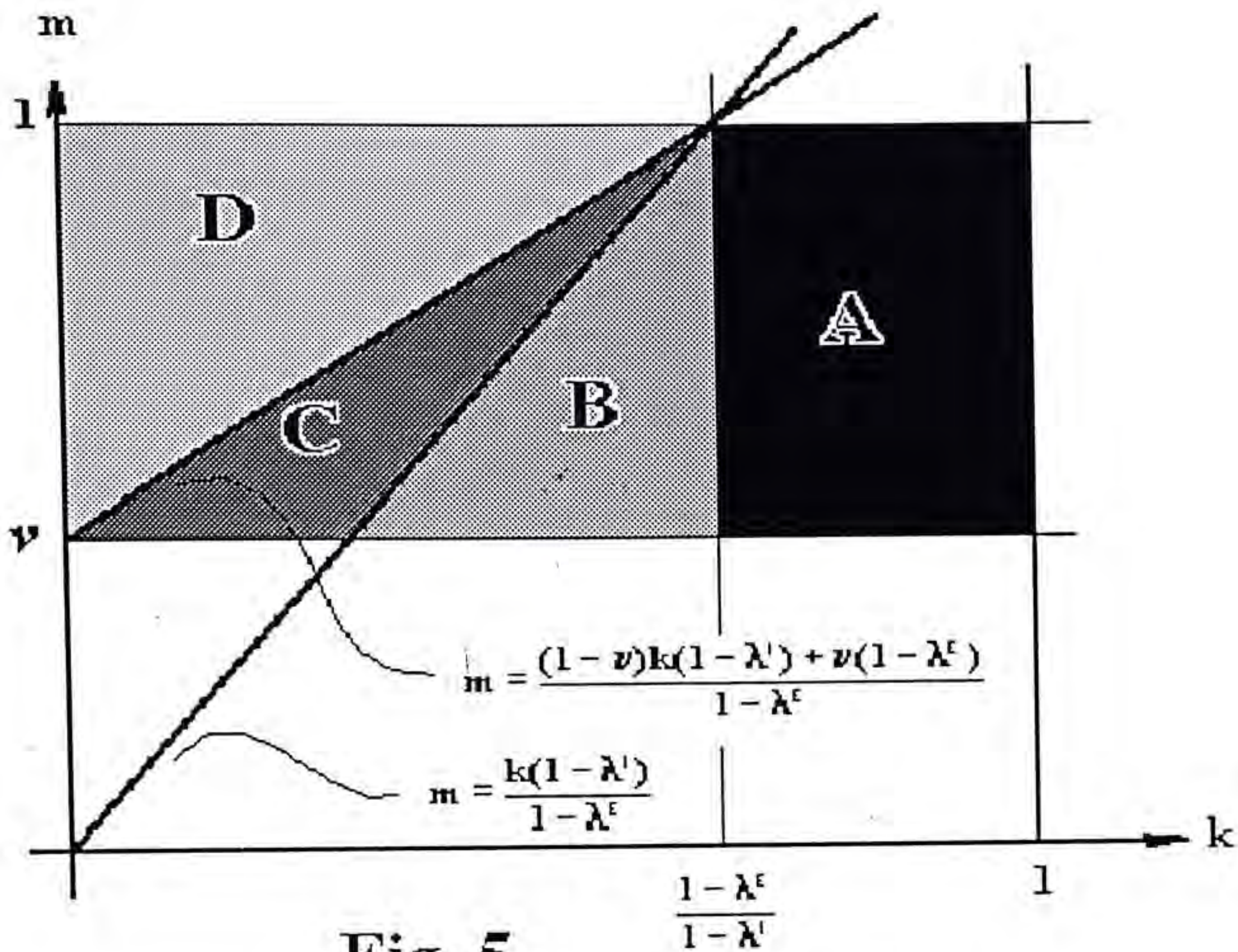

Fig. 5

Proposition 2.3 Under incomplete information with type II informer, if the combination of m and k falls into area A in figure 5, the optimal contract will be characterized by $Q(\omega_1) = Q(\omega_2) = 0$, $\bar{b}_1 = \bar{b}_2 = 1$, $\underline{b}_1 = \underline{b}_2 = \underline{b}(\nu)$, and $\bar{a}_1$, $\bar{a}_2$, $\underline{a}_1$ and $\underline{a}_2$ being given by (19), (20), (25) and (26).

Let's come back to the case where (29) holds. If we assume that (29) holds, the immediate result is that (25) must be binding (or else the central government should keep on reducing both $\bar{a}_1$ and $\underline{a}_1$). It can then be shown by contradiction that Δ must be non-negative.

Lemma 2.7 Δ is non-negative whenever (29) holds.

Proof Suppose not, then there would be no need to reward reporting ω_1, and reducing $\bar{a}_1$ by ε small enough can only benefit the central government. An immediate result is that (19) must be binding.

Under such circumstance, the central government can always enhance its welfare by doing the following:

First, if $\underline{b}_2 > 1$, reduce $\underline{b}_2$ to unity, increase[34] $\underline{a}_2$ and decrease[35] $\bar{a}_2$ such that (20) and (26) remain binding. By doing so, the central government can pay less ($\bar{a}_2$ is decreased) to the productive enterprise, while making the unproductive enterprise produce at the first best level ($\underline{y}(1) = \underline{y}^*$) at the same time. After the change, the payoff of the productive enterprise when the report is ω_2 is still lower than that when the report is ω_1, so $Q(\omega_1)$ is still equal to zero, and the incentive schemes corresponding to report of ω_2 are not affected.

Second, if $\bar{b}_2 \neq 1$, replace $\bar{b}_2$ by unity, adjust $\bar{a}_2$ such that (20) remains binding. Since $\underline{b}_2 \leq 1$, (22') still holds, and the incentive scheme for the unproductive enterprise under ω_2-report is unaffected. As long as (26) holds, the payoff of the productive enterprise under ω_2-report depends only on the incentive scheme for the unproductive one under the same report; as the latter does not change, the former will not change. This in turn guarantees that Δ remains unchanged and $Q(\omega_1)$ remains zero. As a result, by doing so, the central government can gain from making the productive enterprise under ω_2-report produce at the first best level ($\bar{y}(1) = \bar{y}^*$) without the need to pay a penny more.

Third, if $\underline{b}_1 > 1$, reduce $\underline{b}_1$ to unity, increase $\underline{a}_1$ and decrease $\bar{a}_1$ such that (19) and (25) remain binding. As $\bar{a}_1$ decreases, Δ will increase. So it may happen that before $\underline{b}_1$ reaches unity, Δ has already risen to zero. If it is really the case, then just stop there -- we have arrived at a place where Δ is non-negative and the central government is better off than it is initially, and the proof is finished. If it is not the case, we can move on to the next step.

Fourth, if $\bar{b}_1 \neq 1$, replace $\bar{b}_1$ by unity, adjust $\bar{a}_1$ such that (19) remains binding. With similar reason to that in the second step, the central government can gain from making the productive enterprise under ω_1-report produce at the first best level without the need to pay a penny more.

Now we have both $\bar{b}_1$ and $\bar{b}_2$ equal to unity. There can be two mutually exclusive cases: Either $\underline{b}_1 \leq \underline{b}_2$ or $\underline{b}_1 > \underline{b}_2$. The former case implies that $\bar{a}_1 \leq \bar{a}_2$, which in turn implies

---

34 Holding (26) binding, $\frac{d\underline{a}_2}{d\underline{b}_2} = -\underline{y}(\underline{b}_2) < 0$.

35 Holding (20) and (26) binding, $\frac{\partial \bar{a}_2}{\partial \underline{b}_2} = -\underline{y}(\underline{b}_2) + \bar{y}(\underline{b}_2) > 0$

that Δ is non-negative -- a contradiction. Our remaining task is to prove that the latter case can never be optimal.

Recall that the posterior probability of the enterprise to be productive when the informer reports ω_1 is m, and that when the report is ω_2 is $\frac{\nu - nm}{1-n}$, which is smaller than m whenever $m > \nu$. By lemma 2.1, we have $\underline{b}(m) < \underline{b}\left(\frac{\nu - nm}{1-n}\right)$. If $\underline{b}_1 > \underline{b}_2$, either $\underline{b}_1 > \underline{b}(m)$ or $\underline{b}_2 < \underline{b}\left(\frac{\nu - nm}{1-n}\right)$ or both. Then the central government can always gain by moving $\underline{b}_1$ towards $\underline{b}(m)$ and $\underline{b}_2$ towards $\underline{b}\left(\frac{\nu - nm}{1-n}\right)$, until $\underline{b}_1 = \underline{b}_2$.[36] Again we reach a place where Δ is non-negative and the central government is better off than it is initially. **Q.E.D.**

We now examine under what condition will (19) be binding.

Lemma 2.8 (19) will be binding if:

(31) $$m(1-\lambda^E) \geq (1-\lambda^I)k.$$

Proof Suppose not, then lowering $\bar{a}_1$ alone by ε small enough to hold all (binding or unbinding) constraints intact will reduce the expected payment to the enterprise by $nm\varepsilon$ and increase[37] the expected payment to the mid-tier governmental unit by $nk\varepsilon$. The net gain of the central government from doing so is $n\left[m(1-\lambda^E)-(1-\lambda^I)k\right]\varepsilon$, which is independent of the initial level of $\bar{a}_1$. The central government will do so if and only if (31) holds. This means that the initial contract is not yet optimal. **Q.E.D.**

Lemma 2.9 $\Delta = 0$ if the combination of m and k falls in area B.

Proof Suppose not, then the central government can gain by raising $\bar{a}_1$ by ε and lowering $Q(\omega_1)$ by $k\varepsilon$. **Q.E.D.**

$\Delta = 0$ implies (30). So by logic similar to lemma 2.5 and 2.6, we can concluded that:

---

36 During the process $\underline{b}_1$ remains larger than $\underline{b}_2$, $\bar{a}_1$ remains larger than $\bar{a}_2$, and $Q(\omega_1)$ remains zero. The effects on the central government's welfare of the changes in the slopes of the incentive schemes offered to the unproductive enterprise under different informer reports can therefore be viewed separately. As in section three, the central government's welfare is concave in both $\underline{b}_1$ and $\underline{b}_2$, so moving them towards their individual second best levels enhances the central government's welfare.

37 It is because (31) implies (29), which in turn implies that $\Delta \geq 0$.

Proposition 2.4 Under incomplete information with type II informer, if the combination of m and k falls in area B in figure 5, the informer will not be called into service.

Proof Similar to those of lemmas 2.5 and 2.6. **Q.E.D.**

Now if (31) holds, the combination of m and k will fall in areas C or D in figure 5, and constraint (19) will be binding. Assume that $\Delta > 0$, the central government's problem is then to maximize (18) subject to binding constraints (19), (20), (25), (26), $Q(\omega_1) = \Delta$ and (28).[38] Substituting the constraints into the objective function, we can get the following four first order conditions:

$$\frac{\partial W}{\partial \bar{b}_1} = 0 \Rightarrow nm\Big[\bar{y}'(\bar{b}_1) - \lambda^E \bar{\Psi}'(\bar{y}(\bar{b}_1))\bar{y}'(\bar{b}_1) - (1-\lambda^E)(\bar{\Psi}'(\bar{y}(\bar{b}_1))\bar{y}'(\bar{b}_1)) - (1-\lambda^I)k \cdot 0\Big] = 0$$

$$\Rightarrow \bar{y}'(\bar{b}_1) - \lambda^E \bar{b}_1 \bar{y}'(\bar{b}_1) - (1-\lambda^E)\bar{b}_1 \bar{y}'(\bar{b}_1) = 0$$

(32) $$\Rightarrow \bar{b}_1 = 1,$$

$$\frac{\partial W}{\partial \bar{b}_2} = 0 \Rightarrow nm\big[-(1-\lambda^I)k \cdot 0\big] + n(1-m)\cdot 0 + (v - nm)\Big[\bar{y}'(\bar{b}_2) - \lambda^E \bar{b}_2 \bar{y}'(\bar{b}_2) - (1-\lambda^E)(\bar{b}_2 \bar{y}'(\bar{b}_2))\Big] + \big[1 - v - n(1-m)\big]\cdot 0 = 0$$

(33) $$\Rightarrow \bar{b}_2 = 1,$$

$$\frac{\partial W}{\partial \underline{b}_1} = 0 \Rightarrow nm\Big[-(1-\lambda^E)\big(\underline{b}_1 \underline{y}'(\underline{b}_1) - \underline{y}_1(\underline{b}_1) - \underline{b}_1 \underline{y}'(\underline{b}_1) + \bar{y}(\underline{b}_1) + \underline{b}_1 \bar{y}'(\underline{b}_1) - \underline{b}_1 \bar{y}'(\underline{b}_1)\big) - (1-\lambda^I)(-k)\big(\underline{b}_1 \underline{y}'(\underline{b}_1) - \underline{y}(\underline{b}_1) - \underline{b}_1 \underline{y}'(\underline{b}_1) + \bar{y}(\underline{b}_1) + \underline{b}_1 \bar{y}'(\underline{b}_1) - \underline{b}_1 \bar{y}'(\underline{b}_1)\big)\Big]$$
$$+ n(1-m)\Big[\underline{y}'(\underline{b}_1) - \lambda^E \underline{b}_1 \underline{y}'(\underline{b}_1) - (1-\lambda^E)\underline{b}_1 \underline{y}'(\underline{b}_1) - (1-\lambda^I)(-k)\big(\bar{y}(\underline{b}_1) - \underline{y}(\underline{b}_1)\big)\Big] = 0$$

$$\Rightarrow \big(\bar{y}(\underline{b}_1) - \underline{y}(\underline{b}_1)\big)\big[nm(1-\lambda^I)k - nm(1-\lambda^E) + (1-\lambda^I)k(n - nm)\big] + n(1-m)\big[\underline{y}'(\underline{b}_1) - \underline{b}_1 \underline{y}'(\underline{b}_1)\big] = 0$$

(34) $$\Rightarrow 1 - \underline{b}_1 = \frac{\big[m(1-\lambda^E) - k(1-\lambda^I)\big]\big(\bar{y}(\underline{b}_1) - \underline{y}(\underline{b}_1)\big)}{(1-m)\underline{y}'(\underline{b}_1)} > 0, \text{ and}$$

$$\frac{\partial W}{\partial \underline{b}_2} = 0 \Rightarrow nm\Big[-(1-\lambda^I)k\big(-\underline{y}(\underline{b}_2) + \bar{y}(\underline{b}_2)\big)\Big]$$

---

38 We will check later that constraints (21') and (22') are both unbinding.

$$+\,n(1-m)\left[-(1-\lambda^I)k\left(\bar{y}(\underline{b}_2)-\underline{y}(\underline{b}_2)\right)\right]$$
$$+\,(\nu-nm)\left[-(1-\lambda^E)\left(-\underline{y}(\underline{b}_2)+\bar{y}(\underline{b}_2)\right)\right]$$
$$+\left(1-\nu-n(1-m)\right)\left[\underline{y}'(\underline{b}_2)-\lambda^E\underline{b}_2\underline{y}'(\underline{b}_2)\right.$$
$$\left.-(1-\lambda^E)\underline{b}_2\underline{y}'(\underline{b}_2)\right]=0$$
$$\Rightarrow\left(\bar{y}(\underline{b}_2)-\underline{y}(\underline{b}_2)\right)\left[-n(1-\lambda^I)k-(1-\lambda^E)(\nu-nm)\right]$$
$$+\left(1-\nu-n(1-m)\right)(1-\underline{b}_2)\underline{y}'(\underline{b}_2)=0$$

(35)
$$\Rightarrow 1-\underline{b}_2=\frac{\left[n(1-\lambda^I)k+(\nu-nm)(1-\lambda^E)\right]\left(\bar{y}(\underline{b}_2)-\underline{y}(\underline{b}_2)\right)}{\left(1-\nu-n(1-m)\right)\underline{y}'(\underline{b}_2)}>0.$$

The first thing we should note is that (21') and (22') are automatically fulfilled. The second is that Δ is positive if and only if $m>\dfrac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}$.

Lemma 2.10 If (31) holds, Δ is positive if and only if:

(36)
$$m>\frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}.$$

Proof Δ is positive if and only if $\underline{b}_1<\underline{b}_2$. Rewrite (34) and (35) as:

(34')
$$1-\underline{b}_1=\alpha_1\frac{\left(\bar{y}(\underline{b}_1)-\underline{y}(\underline{b}_1)\right)}{\underline{y}'(\underline{b}_1)},\text{ and}$$

(35')
$$1-\underline{b}_2=\alpha_2\frac{\left(\bar{y}(\underline{b}_2)-\underline{y}(\underline{b}_2)\right)}{\underline{y}'(\underline{b}_2)},$$

respectively, where $\alpha_1=\dfrac{m(1-\lambda^E)-k(1-\lambda^I)}{1-m}$ and $\alpha_2=\dfrac{n(1-\lambda^I)k+(\nu-nm)(1-\lambda^E)}{1-\nu-n(1-m)}$.

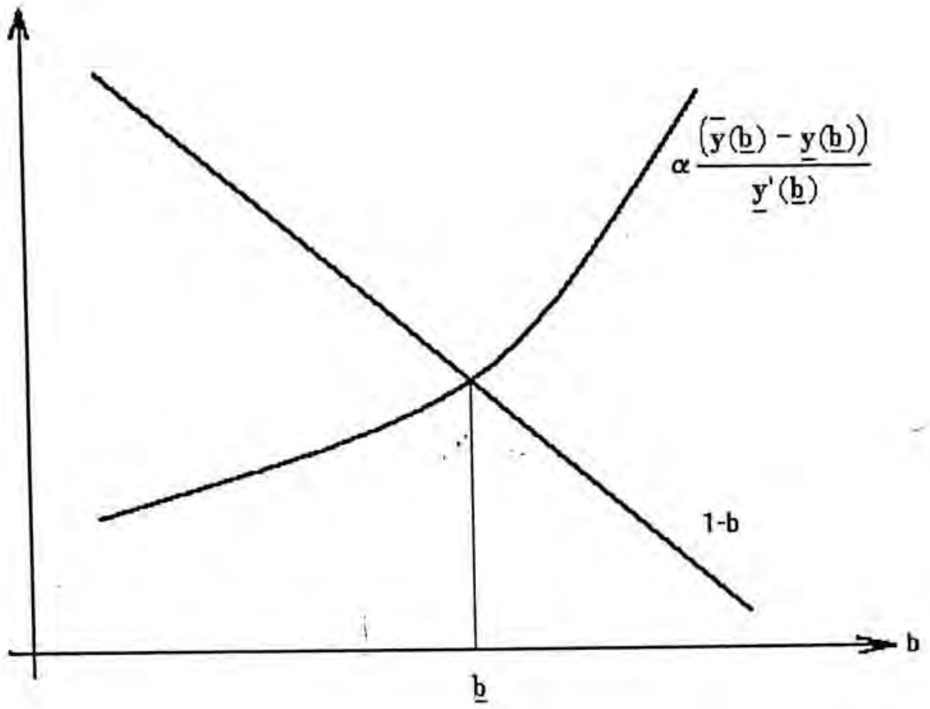

Fig. 6

From figure 6 above, it is clear a larger α will shift up the $\alpha \frac{(\bar{y}(\underline{b}) - \underline{y}(\underline{b}))}{\underline{y}'(\underline{b})}$ schedule and entail in a smaller equilibrium $\underline{b}$. Therefore $\underline{b}_1 < \underline{b}_2$ if and only if $\alpha_1 > \alpha_2$. So it is suffice for us to prove that $\alpha_1 > \alpha_2$ if and only if (36) holds:

$$\frac{m(1-\lambda^E) - k(1-\lambda^I)}{1-m} > \frac{n(1-\lambda^I)k + (\nu - nm)(1-\lambda^E)}{1-\nu-n(1-m)}$$

$$\Leftrightarrow (1-\nu-n(1-m))m(1-\lambda^I) > (1-m)(\nu-nm)(1-\lambda^E)$$

$$\Leftrightarrow m(1-\lambda^E) - (1-\nu)k(1-\lambda^I) > \nu(1-\lambda^E)$$

$$\Leftrightarrow (m-\nu)(1-\lambda^E) > (1-\nu)k(1-\lambda^I)$$

$$\Leftrightarrow m > \frac{(1-\nu)k(1-\lambda^I) + \nu(1-\lambda^E)}{1-\lambda^E}. \qquad \textbf{Q.E.D.}$$

Note that $\frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}$ is in fact a weighted average of $\frac{k(1-\lambda^I)}{1-\lambda^E}$ and unity. If (29) holds, as we have assumed, we will have $\frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E} > \frac{k(1-\lambda^I)}{1-\lambda^E}$. So (36) is a subset of (31). Since (32) - (35) are derived with the assumption that $\Delta > 0$, lemma 2.6 means that this assumption is not self-defeating only in the range where (36) holds. Therefore we have:

Proposition 2.5 Under incomplete information with type II informer, if the combination of m and k falls in area D in figure 5, the optimal contract is characterized by $\bar{b}_1 = \bar{b}_2 = 1$, $\underline{b}_1$ and $\underline{b}_2$ being given by (34) and (35) respectively, $\bar{a}_1$, $\bar{a}_2$, $\underline{a}_1$ and $\underline{a}_2$ being given by (19), (20), (25) and (26) respectively, $Q(\omega_1) = k\Delta$ and $Q(\omega_2) = 0$.

Actually what happens when $m > \frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}$? We can obtain some ideas about it by doing some simple comparative statics. It is easy to see that α_1 is decreasing while α_2 is increasing in k, which implies that $\underline{b}_1$ is increasing and $\underline{b}_2$ is decreasing in k. When k equals zero, that is when the informer cannot receive any bribery at all, $\alpha_1 = \frac{m(1-\lambda^E)}{1-m}$ and $\alpha_2 = \frac{\frac{\nu - nm}{1-n}(1-\lambda^E)}{1-\frac{\nu-nm}{1-n}}$. The resulting $\underline{b}_1$ and $\underline{b}_2$ are therefore $\underline{b}(m)$ and $\underline{b}\left(\frac{\nu-nm}{1-n}\right)$ respectively, where m and $\frac{\nu-nm}{1-n}$ are the posterior probability of an enterprise to be productive when the informer report is ω_1 and ω_2 respectively. It means that if the informer cannot take bribery, the central government can do its best by simply applying the optimal direct mechanism given the posterior probability of an enterprise to be productive conditional upon the informer report.

When k increases from zero, $\underline{b}_1$ increases and $\underline{b}_2$ decreases. The central government finds it desirable to narrow the gap between $\underline{b}_1$ and $\underline{b}_2$ so that it does not need to pay so much to motivate the informer to report truly. As the ability of the informer to take bribery increase, the central government has to narrow the gap more.

According to lemma 2.7, Δ must be non-negative. Combining with lemma 2.9, we therefore have the following lemma:

Lemma 2.11 In the case that $\frac{k(1-\lambda^I)}{1-\lambda^E} < m \le \frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}$, that is when the combination of m and k falls in area C in figure 5, we have $\Delta = 0$.

Since $\Delta = 0$ implies (30), we conclude that:

Proposition 2.6 Under incomplete information with type II informer, if the combination of m and k falls in area C in figure 5, the informer will not be called into service.

Proof Similar to those of lemmas 2.5 and 2.6. **Q.E.D.**

The gap between $\underline{b}_1$ and $\underline{b}_2$ eventually vanishes when k reaches the point such that $(1-\nu)k(1-\lambda^I)=(m-\nu)(1-\lambda^E)$. α_1 and α_2 converge to $\frac{\nu(1-\lambda^E)}{1-\nu}$, while $\underline{b}_1$ and $\underline{b}_2$ converge to $\underline{b}(\nu)$. Now the incentive schemes offered to the productive and unproductive enterprises become the same under the two different reports. In other words, the central government will do exactly what it will do with a type I informer or without informer at all.

As a summary of propositions 2.3 - 2.6, an informer will be called into service only if $m>\frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}$. That is, only when the signal observed by the informer is Ω-informative enough. When the informer is too Ω-uninformative, the central government will find the benefit brought by the signal too small to cover the cost incurred in motivating the informer to tell the truth, and the informer will therefore not be called into service.

Section six: Incomplete information with type III informer

A type III informer is one whose verifiability structure is $\{\phi,\{\omega_2\},\Omega\}$, and therefore can verify ω_2 but not ω_1. So it will be bribed to mis-report ω_1 when it actually observes ω_2, but not vice versa. In order to motivate the informer to report truly, the central government has to reward reporting ω_2. The minimum reward is the maximum bribery an informer observing ω_2 can receive if it reports ω_1. Thus we have the following modified informer incentive compatibility constraints:

(27") $$Q(\omega_2)-Q(\omega_1)\geq k\cdot\max\{[\overline{R}_1-\overline{\Psi}(\overline{y}(\overline{b}_1))]-[\overline{R}_2-\overline{\Psi}(\overline{y}(\overline{b}_2))],\ [\underline{R}_1-\underline{\Psi}(\underline{y}(\underline{b}_1))]-[\underline{R}_2-\underline{\Psi}(\underline{y}(\underline{b}_2))],0\},$$ and

(28') $$Q(\omega_1)\geq 0.$$

The central government's problem is therefore to maximize (18) subject to constraints (19) - (26), (27") and (28'). The main features of the optimal contract is summarized by the following proposition.

Proposition 2.7 Under incomplete information with type III informer, the optimal contract is characterized by $Q(\omega_1)=Q(\omega_2)=0$, $\overline{b}_1=\overline{b}_2=1$, $\underline{b}_1=\underline{b}(m)$, $\underline{b}_2=\underline{b}\left(\frac{\nu-nm}{1-n}\right)$, and $\overline{a}_1$, $\overline{a}_2$, $\underline{a}_1$ and $\underline{a}_2$ being given by (19), (20), (25) and (26) respectively.

Section seven: Incomplete information with type IV informer

As argued in chapter one, a type IV informer, whose verifiability structure is $\{\phi, \{\omega_1\}, \{\omega_2\}, \Omega\}$, cannot cheat the central government. So after paying the reservation wage to the informer, the central government can take over all the information the informer has. The problem of the central government is therefore to maximize (18) subject to (19) - (26). The optimal contract is given by the following proposition.

Proposition 2.8 Under incomplete information with type IV informer, the optimal contract is characterized by $Q(\omega_1) = Q(\omega_2) = 0$, $\bar{b}_1 = \bar{b}_2 = 1$, $\underline{b}_1 = \underline{b}(m)$, $\underline{b}_2 = \underline{b}\left(\frac{\nu - nm}{1-n}\right)$, and $\bar{a}_1$, $\bar{a}_2$, $\underline{a}_1$ and $\underline{a}_2$ being given by (19), (20), (25) and (26) respectively.

Section eight: Comparative values of mid-tier governmental units to the central government

The following results are immediate from propositions 2.2 - 2.8.

Proposition 2.9 For mid-tier governmental units of the same type, the lower is the rank of a mid-tier governmental unit in the hierarchy, the more valuable is the mid-tier governmental unit to the central government.

Proof The proof for types I informers is immediate. The central government will consider all type I informers as non-valuable at all, and so every type I informer is at least as valuable than other type I informers. The same is true for type II informers when $m \leq \frac{(1-\nu)k(1-\lambda^I) + \nu(1-\lambda^E)}{1-\lambda^E}$. The proof for type III and IV informers is also straight forward. The central government will take over all the information hold by a type III or IV informer, so the value of the informer is positively linked to the Ω-informativeness of the signal he manages to observe. As low-ranked governmental units can observe more Ω-informative signals, they are therefore more valuable. So what we need to prove is that the proposition is true for type II informers when $m > \frac{(1-\nu)k(1-\lambda^I) + \nu(1-\lambda^E)}{1-\lambda^E}$.

Denote by W* the maximized expected welfare of the central government when the informer is of type II. It suffices to show that the first derivatives of W* with respect to m and n are both non-negative. Differentiating W* with respect to m, we have:

$$\frac{\partial W^*}{\partial m} = n\Big\{\Big[\bar{y}(\bar{b}_1) - \lambda^E\bar{\Psi}(\bar{y}(\bar{b}_1)) - (1-\lambda^E)\bar{R}_1\Big] - \big(\underline{y}(\underline{b}_1) - \underline{R}_1\big)$$
$$-\Big[\bar{y}(\bar{b}_2) - \lambda^E\bar{\Psi}(\bar{y}(\bar{b}_2)) - (1-\lambda^E)\bar{R}_2\Big] + \big(\underline{y}(\underline{b}_2) - \underline{R}_2\big)\Big\}$$
$$= n\Big\{(1-\lambda^E)(\bar{a}_2 - \bar{a}_1) - \big[\underline{y}(\underline{b}_1) - \underline{\Psi}(\underline{y}(\underline{b}_1))\big] + \big[\underline{y}(\underline{b}_2) - \underline{\Psi}(\underline{y}(\underline{b}_2))\big]\Big\},$$

which is non-negative. Similarly, differentiating W* with respect to n, we have:

$$\frac{\partial W^*}{\partial n} = m\Big[\bar{y}(\bar{b}_1) - \lambda^E\bar{\Psi}(\bar{y}(\bar{b}_1)) - (1-\lambda^E)\bar{R}_1 - (1-\lambda^I)Q(\omega_1)\Big]$$
$$+(1-m)\Big[\underline{y}(\underline{b}_1) - \lambda^E\underline{\Psi}(\underline{y}(\underline{b}_1)) - (1-\lambda^E)\underline{R}_1 - (1-\lambda^I)Q(\omega_1)\Big]$$
$$-m\Big[\bar{y}(\bar{b}_2) - \lambda^E\bar{\Psi}(\bar{y}(\bar{b}_2)) - (1-\lambda^E)\bar{R}_2\Big]$$
$$-(1-m)\Big[\underline{y}(\underline{b}_2) - \lambda^E\underline{\Psi}(\underline{y}(\underline{b}_2)) - (1-\lambda^E)\underline{R}_2\Big]$$
$$= \big((1-\lambda^E)m - (1-\lambda^I)k\big)(\bar{a}_2 - \bar{a}_1) - (1-m)(\underline{a}_2 - \underline{a}_1).$$

The first term in the last expression is positive, while the second one is negative. This makes the sign of the last expression not immediately clear. To work out the sign of $\frac{\partial W^*}{\partial n}$, we first derive the second derivative of W* with respect to n:

$$\frac{\partial^2 W^*}{\partial n^2} = \big((1-\lambda^E)m - (1-\lambda^I)k\big)\frac{\partial \bar{a}_2}{\partial n} - (1-m)(1-\underline{b}_2)\frac{\partial \underline{b}_2}{\partial n}$$
$$= (1-m)\underline{y}'(\underline{b}_2)\left[\frac{\big((1-\lambda^E)m - (1-\lambda^I)k\big)\big(\bar{y}(\underline{b}_2) - \underline{y}(\underline{b}_2)\big)}{(1-m)\underline{y}'(\underline{b}_2)} - (1-\underline{b}_2)\right]\frac{\partial \underline{b}_2}{\partial n}.$$

The derivatives of $\bar{a}_1$ and $\underline{b}_1$ with respect to n vanish because $\underline{b}_1$ is independent of n. The expression in the bracket equals to the length $ac - ab$ in figure 7, and is thus positive. $\frac{\partial \underline{b}_2}{\partial n}$ is positive provided $m > \frac{(1-\nu)k(1-\lambda^I) + \nu(1-\lambda^E)}{1-\lambda^E}$. So we have the second derivative of W* with respect to n positive.

As the derivative of W* with respect to n is increasing in n, it becomes suffice for us to show that it is positive when n approaches its lower bound, that is zero. Denote $\lim_{n\to 0}\underline{b}_2$ by $\underline{b}$, and $\lim_{n\to 0}\bar{a}_2$ by $\bar{a}$. We have:

$$\lim_{n\to 0}\frac{\partial W^*}{\partial n} = (\bar{a} - \bar{a}_1)\big((1-\lambda^E)m - (1-\lambda^I)k\big)$$
$$+(1-m)\big[\underline{y}(\underline{b}_1) - \underline{\Psi}(\underline{y}(\underline{b}_1))\big] - (1-m)\big[\underline{y}(\underline{b}) - \underline{\Psi}(\underline{y}(\underline{b}))\big].$$

Denote the last expression by $\Phi(\underline{b}_1)$. Since $\frac{\partial \Phi}{\partial \underline{b}_1} = \frac{\partial}{\partial \underline{b}_1}\left(\frac{\partial W^*}{\partial n}\right) = \frac{\partial}{\partial n}\left(\frac{\partial W^*}{\partial \underline{b}_1}\right) = 0$, and

$$\frac{\partial^2 \Phi}{\partial \underline{b}_1^{\,2}} = \big((1-\lambda^E)m - (1-\lambda^I)k\big)(-1)\big(\bar{y}'(\underline{b}_1) - \underline{y}'(\underline{b}_1)\big)$$

$$+(1-m)(-1)\underline{y}'(\underline{b}_1)+(1-m)(1-\underline{b}_1)\underline{y}''(\underline{b}_1)<0,$$

we have $\Phi(\underline{b}_1) > \Phi(\underline{b}) = 0$. Therefore we have $\lim_{n\to 0}\frac{\partial W^*}{\partial n} > 0$, and hence $\frac{\partial W^*}{\partial n} > 0$ for all n in the interval [0 , 1]. **Q.E.D.**

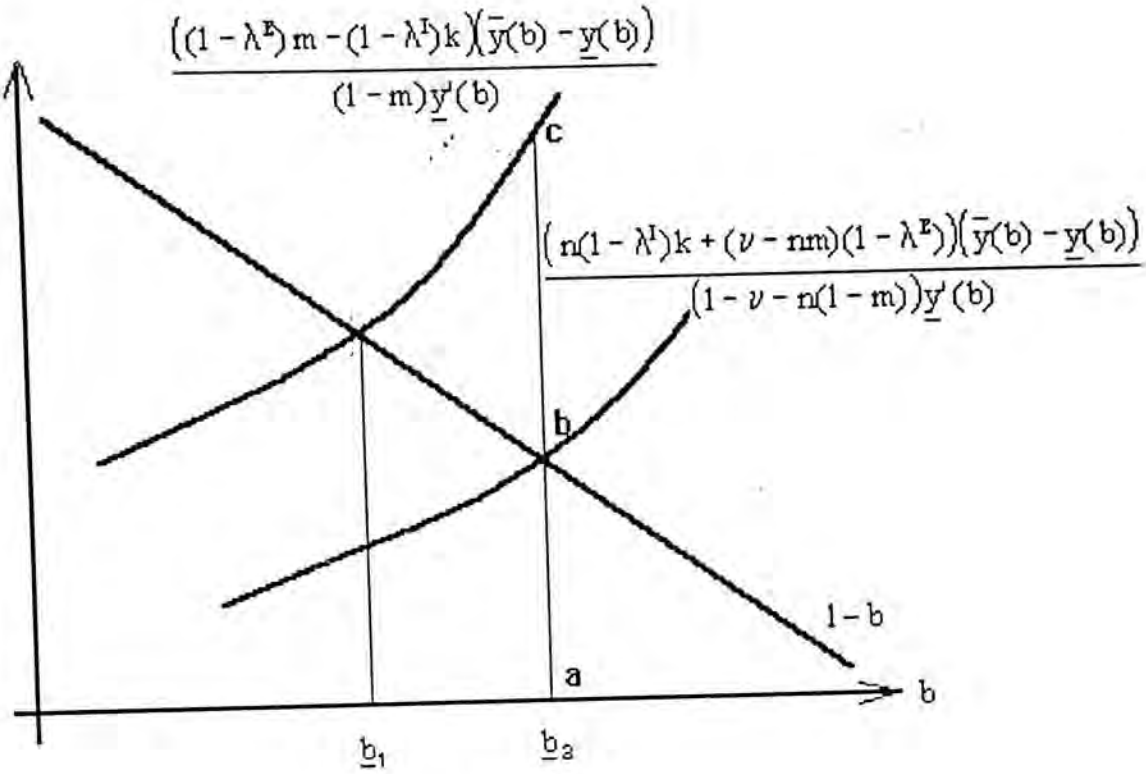

Fig. 7

Proposition 2.9 is a restatement of theorem 1.2 in chapter one. It suggests a possible reason why people sometimes have such illusion that idiosyncratic responsibility contracts which utilize more local information can better enhance the centre's payoff. It also tells us what assumption they have implicitly made when they claim so. The following proposition says that when this assumption does not hold, things will be very different.

Proposition 2.10 A low-ranked governmental unit is less valuable to the central government than a high-ranked governmental unit if one of the following is true:

(i) the low-ranked unit is of type I while the high-ranked unit is of type III or IV; or

(ii) the low-ranked unit is of type I while the high-ranked unit is of type II with $m > \frac{(1-\nu)k(1-\lambda^I)+\nu(1-\lambda^E)}{1-\lambda^E}$; or

(iii) the low-ranked unit is of type II with $m \leq \dfrac{(1-\nu)k(1-\lambda^{I})+\nu(1-\lambda^{E})}{1-\lambda^{E}}$, while the high-ranked unit is of type III or IV; or

(iv) the low-ranked unit is of type II with $m > \dfrac{(1-\nu)k(1-\lambda^{I})+\nu(1-\lambda^{E})}{1-\lambda^{E}}$, while the high-ranked unit is of type III or IV, and their ranks do not differ too much from each other.

Proof The proofs for (i) - (iii) are immediate. To prove (iv), it is suffice to prove that when $m > \dfrac{(1-\nu)k(1-\lambda^{I})+\nu(1-\lambda^{E})}{1-\lambda^{E}}$,

$$W_{III}^*(m,n) = W_{IV}^*(m,n) > W_{II}^*(m,n), \tag{37}$$

where $W_{III}^*(m,n)$, $W_{IV}^*(m,n)$ and $W_{II}^*(m,n)$ are the central government's maximized expected payoffs when the mid-tier governmental unit is of types III, IV and II, respectively. Since $W_{II}^*(m,n)$ is continuous in m and n, (37) will imply that there exists $(m',n') >> (m,n)$ such that $W_{III}^*(m,n) = W_{IV}^*(m,n) > W_{II}^*(m',n') > W_{II}^*(m,n)$, and (iv) will then follow.

To prove (37), note that as the problem of the central government has less constraints when the mid-tier governmental unit is of type IV than when it is of type II. So we already have $W_{III}^*(m,n) = W_{IV}^*(m,n) \geq W_{II}^*(m,n)$. Our task is therefore to show that the inequality sign is in fact a strict one.

Denote the vector $(\underline{b}_1, \underline{b}_2)$ when the informer is of type II by $(\hat{b}_1, \hat{b}_2)$, and that when the informer is of type III or IV by $(\tilde{b}_1, \tilde{b}_2)$. Note that we have already show that $\hat{b}_1 > \tilde{b}_1$ and $\hat{b}_2 < \tilde{b}_2$ in section five (see pp.44-45). Denote by $\Gamma(\underline{b}_1, \underline{b}_2)$ the following expression:

$$\begin{aligned}
&nm\left[\overline{y}(\overline{b}_1) - \lambda^{E}\overline{\Psi}(\overline{y}(\overline{b}_1)) - (1-\lambda^{E})\overline{R}_1\right] \\
&+n(1-m)\left[\underline{y}(\underline{b}_1) - \lambda^{E}\underline{\Psi}(\underline{y}(\underline{b}_1)) - (1-\lambda^{E})\underline{R}_1\right] \\
&+(\nu-nm)\left[\overline{y}(\overline{b}_2) - \lambda^{E}\overline{\Psi}(\overline{y}(\overline{b}_2)) - (1-\lambda^{E})\overline{R}_2\right] \\
&+(1-\nu-n(1-m))\left[\underline{y}(\underline{b}_2) - \lambda^{E}\underline{\Psi}(\underline{y}(\underline{b}_2)) - (1-\lambda^{E})\underline{R}_2\right].
\end{aligned}$$

Clearly $W_{III}^*(m,n) = W_{IV}^*(m,n) = \Gamma(\tilde{b}_1, \tilde{b}_2)$, and $W_{II}^*(m,n) = \Gamma(\hat{b}_1, \hat{b}_2) - n(1-\lambda^{I})Q(\omega_1)$. So it is suffice for us to show that $\Gamma(\tilde{b}_1, \tilde{b}_2) > \Gamma(\hat{b}_1, \hat{b}_2)$.

Note that $\Gamma(\underline{b}_1, \underline{b}_2)$ is strictly concave in $\underline{b}_1$ and $\underline{b}_2$,[39] and achieves its maximum at $(\tilde{b}_1, \tilde{b}_2)$. Therefore we have $\Gamma(\tilde{b}_1, \tilde{b}_2) > \Gamma(\hat{b}_1, \hat{b}_2)$. **Q.E.D.**

---

39 For a proof see footnote 19.

Proposition 2.10 suggests how fragile is the believe that low-ranked governmental units can always provide more valuable information to the central government.

Section nine: Conclusion

In this chapter we have demonstrated how a mid-tier governmental unit observing a more Ω-informative signal can happen to be less valuable to the central government. The key of such phenomenon lies on the difference in verifiability structures among different mid-tier governmental units. This finding sheds new light on the direction of future research on the centre-enterprise relationship. Field research on how mid-tier governmental units justify the terms of the responsibility contracts they offered to the enterprises under their jurisdiction to the centre must now enter our research agenda.

One of the possible causes for the difference in verifiability structures may lie in the difference in the ways different mid-tier governmental units collect their information about the enterprises under their jurisdiction. For example, high-ranked governmental units often collect their information by undertaking nationwide *tuji jiancha* (sudden inspections), while low-ranked governmental units collect their information through daily interaction with the enterprises. Information collected during the sudden inspections, like stocks of working capital or vintages of physical capital, has the nice property that it can be readily verified, while information collected through daily interaction with the enterprises, like problem-solving ability of the managers or morale of the workers, is mostly non-verifiable. Therefore different mid-tier governmental units not only observe different signals, but can have very different verifiability structure as well.[40]

The result that low-ranked governmental units may not be informationally more valuable than their high-ranked counterparts helps to explain how idiosyncratic responsibility contracts may turn out to be harmful to the centre by making use of the more v-informative but less verifiable signals instead of the more verifiable though less v-informative signals.

---

40 I am indebted to Kai-Yuen Tsui for suggesting this example.

Yet there remains an important gap to be filled if we want to have a complete explanation for the emergence of uniform responsibility contracts. Indeed, what our theory has actually shown is that idiosyncratic responsibility contracts that involve low-ranked governmental units may be informationally inferior to those that involve high-ranked governmental units. Though it is natural to suppose that more uniform responsibility contracts will be concluded when less local information is employed, our theory does not tell why. Worse still, the contract theory does suggest that if it is known that there are n different types of enterprise, the optimal contract should at least contain n different incentive schemes, each to be offered to one type of enterprise. It means if we suppose that almost all enterprises differ to some extent from the others, there should be as many different incentive schemes as the number of enterprises in China. But it is not what we have actually observed. So there must be something missed in our theory.

One of the possible ways to fill this gap is to introduce some costs to the construction of every additional incentive scheme, so that it is not cost-efficient to offer to many incentive schemes. Another way is to admit that the center does not have prior knowledge on the possible types of enterprise. The center only knows that a particular type of enterprise exists not until some mid-tier governmental units report (with verification) so to the center. When the center employs less local information, fewer different incentive schemes will be offered, and the responsibility contracts will become more uniform across different enterprises.

REFERENCE

Arrow, K.J. 1974. "Limited Knowledge and Economic Analysis." *American Economic Review* **64**(1), 1-10.

Arrow, K.J. 1985. "Informational Structure of the Firm." *American Economic Review Papers and Proceedings* 75(2), 303-307.

Arrow, K.J. 1986. "Information and the Market." In K.J. Arrow and M.D. Intriligator (eds.) *Handbook of Mathematical Economics vol. 3*. North-Holland.

Arrow, K.J. 1992. "Informational Equivalence of Signals." In P. Dasgupta, D. Gale, O. Hart, E. Maskin (eds.) *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*. MIT Press.

Blackwell, D. 1951. "Comparison of Experiments." In J. Neymann (ed.) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press.

Chai, J.C.H. 1991. "Incentive and Profit Sharing in Chinese Industry." *Hong Kong Economic Papers* **21**, 47-57.

Cremer, J. 1982. "A Simple Proof of Blackwell's 'Comparison of Experiments' Theorem." *Journal of Economic Theory* **27**, 439-443.

Dong, Q. 1988. *Guojia Shuishou* (The Tax System of the Country). Liaoning: Xibei Caijing Daxue Chubanshe.

Du H. and R. Zhang 1992. "*1979 nian yilai Guoyou Qiye Gaige Huigu yu Zhanwang* (The History and Prospect of Chinese State Enterprise Reform since 1979)." In F. Dong and Z. Tang eds. *Zhongguo Guoyou Qiye Gaige: Zhidu yu Xiaolu* (Chinese State Enterprise Reform: Institution and Efficiency). Beijing: Zhongguo Jihua Chubanshe.

Freixas, X., R. Guesnerie and J. Tirole 1985. "Planning under Incomplete Information and the Ratchet Effect." *The Review of Economic Studies* **52**, 173-191.

Green, J.R. and J-J Laffont. 1986. "Incentive Theory with Data Compression." In W.P. Heller, R.M. Starr and D.A. Starrett (eds.) *Uncertainty, Information, and*

Communication: Essays in Honor of Kenneth J. Arrow, Volume III. Cambridge University Press.

Laffont, J.-J. 1989. *The Economics of Uncertainty and Information.* Translated by J.P. Bonin and H. Bonin. MIT Press.

Marschak, T. 1987. "Organization Theory." In J. Eatwell, M. Milgate and P. Newman (eds.) *The New Palgrave: A Dictionary of Economics.* McMillan Press.

Myerson, R.B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* **47**(1), 61-73.

Oniki, H. 1986. "The Cost of Communication in Economic Organization: II." In W.P. Heller, R.M. Starr and D.A. Starrett (eds.) *Uncertainty, Information, and Communication: Essays in Honor of Kenneth J. Arrow, Volume III.* Cambridge University Press.

Rogerson, W.P. 1985. "The First-Oder Approach to Principal-Agent Problems." *Econometrica* **53**, 1357-1367.

Sah, R.K. and J.E. Stiglitz. 1986. "The Architecture of Economic Systems: Hierarchies and Polyarchies." *American Economic Review* **76**(4), 716-727.

State Statistical Bureau 1991. *Zhongguo Tongji Nianjian 1991* (Chinese Statistical Yearbook 1991).

Tirole, J. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, and Organization* **2**, 181-214.

Tirole, J. 1992. "Collusion and the Theory of Organizations." in J-J Laffont (ed.) *Advances in Economic Theory: Sixth World Congress vol. 2.* New York: Cambridge University Press.

Wong, C.P.W. 1987. "Between Plan and Market: The Role of the Local Sector in Post-Mao China." Journal of Comparative Economics **11**, 385-398.

Wu, J., and J. Liu. 1991. *Lun Jingzhengxing Shichang Tizhi.* Beijing: Zhongguo Caizheng Jingji Chubanshe.